Estimation of Distances between Point Process Distributions

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Dominic Schuhmacher

von

Kloten ZH

Promotionskomitee

Prof. Dr. Andrew Barbour (Vorsitz)
Prof. Dr. Erwin Bolthausen
Prof. Dr. Ilya Molchanov

Zürich, 2005

Abstract

Point processes serve as probabilistic models for random locally finite point patterns on some general space. Their distributions are typically rather involved objects, and it is therefore of considerable interest to approximate the more complicated of these by relatively simple ones, most prominently by the distribution of a Poisson process; this process scatters a Poisson number of points independently and with equal probability laws over any given compact set. There are many theorems available that state convergence in distribution of a sequence of more general point processes towards a Poisson process.

In the present work, concrete distance information is provided for some of these convergence theorems. As our main metric for measuring the corresponding distances between point process distributions, we use the Wasserstein metric d_2, which was introduced in Barbour and Brown (1992) [Stochastic Process. Appl. 43, 9–31] and has since then been applied in several studies.

A first focal point of this thesis is to provide comprehensive information about the d_2-metric with regard to its theoretical properties, as well as its more applied aspects. We partly gather results from various sources in the literature, supplying formal proofs where these were previously missing, and partly add new results that do not seem to be published so far. This topic is covered in Chapter 3 and, to some extent, also in Chapter 4.

The main emphasis of the thesis rests then on evaluating the d_2-distances for settings that essentially correspond to those of certain Poisson process limit theorems. Concrete upper bounds for such distances are given which imply convergence under conditions that are usually only slightly stronger than those in the original limit theorems, and which are, of course, much more informative since they provide explicit convergence rates. Furthermore, these upper bounds remain valid also under more general conditions, where convergence does not necessarily hold. Our main approximation situations concern point processes subjected to certain linear stretch-contract transformations, to random thinnings that are accompanied by contractions of the space, and to superpositions that are accompanied by a mechanism that makes the point processes sparser. These topics are covered in Chapters 5, 6, and 7, respectively.

Most of the distance bounds are derived either directly or indirectly by means of Stein's method, which is a powerful technique for estimating distances between probability distributions in various situations. A brief introduction to Stein's method is presented in Chapter 4.

As a prerequisite for reading this thesis, elementary knowledge of measure theoretically founded probability theory is sufficient. An overview of the more special concepts that are used here is given in Chapter 2.

Zusammenfassung

Punktprozesse dienen als Wahrscheinlichkeitsmodelle für zufällige lokalendliche Punktmuster auf einem allgemeinen Raum. Ihre Verteilungen sind meist ziemlich komplexe Objekte, und es ist daher von beträchtlichem Interesse, die komplizierteren Exemplare unter ihnen durch relativ einfache, am häufigsten durch eine Poissonprozess-Verteilung, zu approximieren. Ein Poissonprozess streut eine poissonverteilte Anzahl Punkte unabhängig voneinander mit derselben Wahrscheinlichkeitsverteilung über jede feste kompakte Menge. Es stehen viele bekannte Theoreme zur Verfügung, die Aussagen machen über die Konvergenz in Verteilung einer Folge von allgemeineren Punktprozessen gegen einen Poissonprozess.

In der vorliegenden Arbeit werden konkrete Distanzinformationen für einige dieser Approximationsprobleme gegeben. Die Metrik, die wir hauptsächlich benutzen, um die entsprechenden Distanzen zwischen Verteilungen von Punktprozessen zu messen, ist die Wassersteinmetrik d_2, die in Barbour und Brown (1992) [Stochastic Process. Appl. 43, 9–31] eingeführt und seit damals in verschiedenen Studien verwendet wurde.

Ein erster Schwerpunkt dieser Doktorarbeit ist die Bereitstellung umfassender Informationen über die d_2-Metrik, sowohl hinsichtlich ihrer theoretischen Eigenschaften, als auch ihrer angewandteren Aspekte. Zum Teil tragen wir Resultate aus der Literatur zusammen und liefern Beweise nach, wo diese bisher gefehlt haben, zum Teil fügen wir neue Resultate hinzu, die bisher nicht veröffentlicht zu sein scheinen. Dies ist der Inhalt von Kapitel 3 und, in geringerem Ausmass, von Kapitel 4.

Das Hauptthema der Doktorarbeit ist dann die konkrete Evaluation der d_2-Distanzen in Situationen, die im wesentlichen denen von gewissen Poissonprozess-Grenzwertsätzen entsprechen. Es werden konkrete obere Schranken für diese Distanzen hergeleitet, die Konvergenz implizieren unter Bedingungen, welche meist nur leicht stärker sind als diejenigen der ursprünglichen Grenzwertsätze, und die natürlich sehr viel aussagekräftiger sind, da sie explizite Konvergenzraten liefern. Hinzu kommt, dass die oberen Schranken unter allgemeineren Bedingungen gültig bleiben, unter denen unter Umständen keine Konvergenz mehr gilt. Unsere hauptsächlichen Approximationsszenarien betreffen Punktprozesse, die entweder durch gewisse lineare Streck-stauch-Transformationen modifiziert werden, oder durch Verdünnungen, die von Kontraktionen des Raumes begleitet sind, oder durch Superpositionen, mit denen ein Verfahren einhergeht, das die Punktprozesse spärlicher werden lässt. Diese Themen werden der Reihe nach in den Kapiteln 5, 6 und 7 behandelt.

Die meisten der Distanzabschätzungen werden auf direkte oder indirekte Weise mittels der Steinschen Methode erhalten. Diese ist ein mächtiges Werkzeug zur Abschätzung von Distanzen zwischen Wahrscheinlichkeitsverteilungen in vielerlei Situationen. Eine kurze Einführung in die Steinsche Methode ist der Gegenstand von Kapitel 4.

Als Voraussetzung zum Lesen dieser Dissertation genügen elementare Kenntnisse in masstheoretisch fundierter Wahrscheinlichkeitstheorie. Ein Überblick über die spezielleren Konzepte, die hier benutzt werden, wird in Kapitel 2 gegeben.

Acknowledgement

Many people have accompanied and supported me along the long and winding path of writing this thesis. Now that I have almost arrived at my destination, it is with great pleasure that I look back and express my gratitude to everyone who has been a map maker, a guide, or a fellow traveller. A warm and heartfelt thank you goes to those who helped out with directions at the crossings, to those who shouted their encouragement or their displeasure from the wayside, as well as to those who offered a cool beer during the march breaks.

There are a few people that I would like to mention explicitly. I am certainly most indebted to my advisor, Professor Andrew Barbour, for his continued support and the many helpful discussions throughout my studies. With his positive attitude and his patience, especially towards many of my rather technical approaches, he created a very pleasant atmosphere and I would like to thank for the many things I was able to learn from him about mathematics and otherwise.

I am especially grateful to Professor Ilya Molchanov for his kind support in connection with the thinning problem. My sincere thanks are furthermore given to Professors Aihua Xia, Tim Brown, and Adrian Baddeley for brief, but very helpful discussions.

Part of the research for the superposition problem was carried out while I was visiting the National University of Singapore, and I would like to thank Professor Louis Chen and the Institute for Mathematical Sciences for their hospitality and support.

I am greatly indebted to many of my colleagues at the Institute of Mathematics for everything from the everyday chitchat during the tea breaks to deep mathematical discussions. My special thanks go to Adrian Röllin, Bruno Nietlispach and Urs Kollbrunner, who have been a constant source of mathematical wisdom and entertainment.

Last, but definitely not least, I would like to express my deep and sincere gratitude to my parents for all and everything that I have come to be.

Dominic Schuhmacher
Zürich, July 2005

Contents

Chapter 1

Introduction

Point processes are the fundamental probabilistic objects for modeling spatial information. They are comparatively easy to handle mathematically, and often successfully describe the essence of spatial real world phenomena.

Mathematically, a point process is defined as a random locally finite point measure on a space E, which is usually assumed to be a locally compact second countable Hausdorff space. Since the space $\mathfrak{N}$ of such point measures inherits many of the nice properties of E, it is possible to deal with point processes formally without much of a technical apparatus. Another advantage is that point processes are very well studied objects. There are comprehensive works available dealing with their theory, such as Kallenberg (1986) and Daley and Vere-Jones (1988), and they are at the center of extensive research into both their theoretical and their applied properties.

Intuitively, a point process can be thought of as a random point pattern on E. In practice, these patterns are used most often to model actual spatial structures in two or three dimensions, or, more traditionally, series of events in time, so that very often $E = \mathbb{R}^D$ with $D \in \{1, 2, 3\}$. Such models are used in many different domains, such as forestry (the points are typically the centers of tree trunks or the positions of seedlings), seismology (the points describe the positions of the epicenters of earthquakes and possibly the times of their occurences), geography (e.g. human settlements or towns), zoology (nests or burrows), astronomy (stars or galaxies), spatial epidemiology (positions of hosts), cell biology (centers of cells or positions of their nuclei), or material sciences (centers of cavities in porous rock, of defects in an industrial material, and so on). A selection of real world point pattern data of a very diverse nature is given in Figure 1.1.

A class of point processes of outstanding importance is that of the Poisson processes. On a compact state space E' such a process is given by scattering a Poisson distributed number of points independently over E', all according to the same arbitrary distribution. This leads to the fundamental property that a Poisson process has independent numbers of points in any two disjoint regions of the state space, which makes it particularly nice to deal with, and is responsible for the fact that a comparatively large number of functionals of Poisson processes can be explicitly evaluated. The mathematical tractability, along with the fact that, in many circumstances, the Poisson process turns out to be a rather realistic model for observed spatial or temporal data, make this kind of process invaluable also for practical purposes.

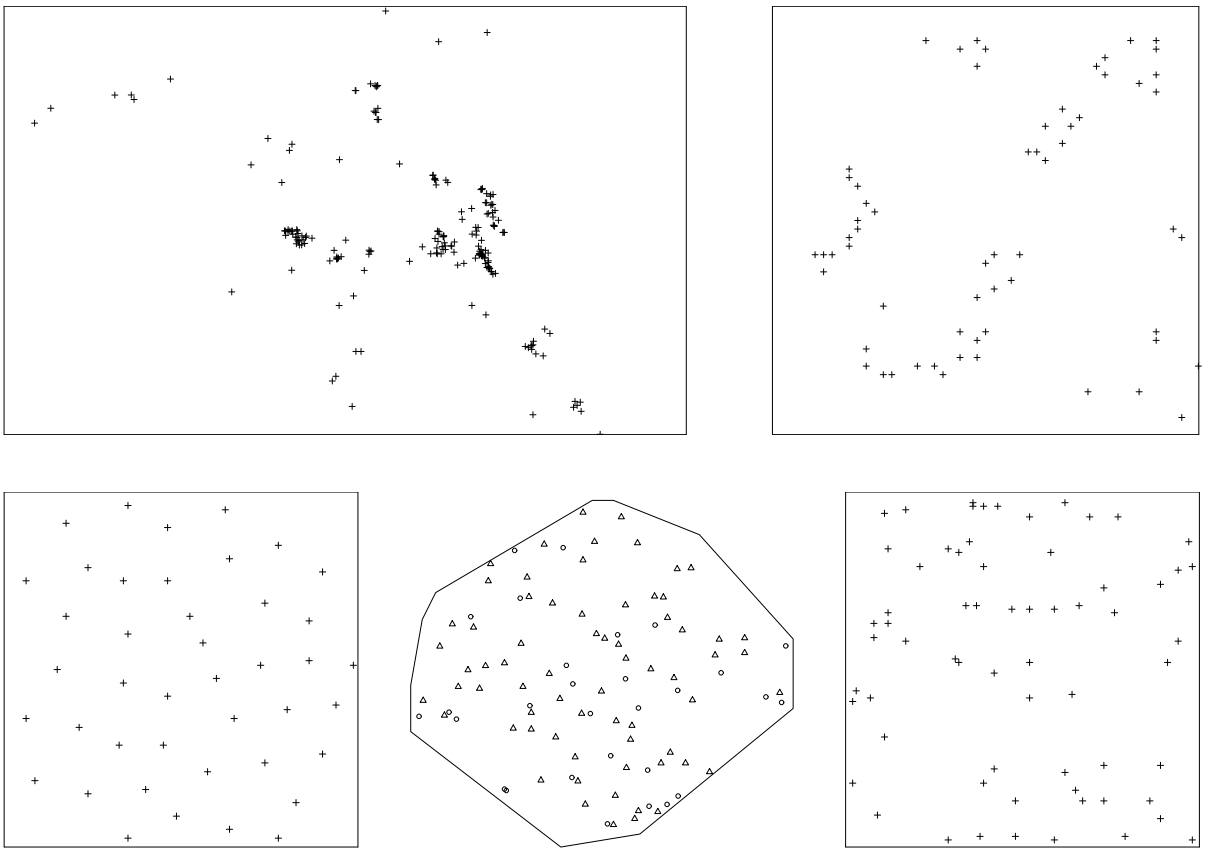

FIGURE 1.1: EXAMPLES OF POINT PATTERN DATA. *Top row:* Epicenters of earthquakes in Southern California; seedlings and saplings of California redwood trees. *Bottom row:* Centers of biological cells; nests of two species of ants in northern Greece; saplings of Japanese black pines. Details for the different data sets are given in Appendix A.5.

Figure 1.2 conveys a feeling for the random mechanism by which a Poisson process scatters its points. On display are six realizations of a homogeneous Poisson process with intensity 70 on the unit square, which means that a number of points that follows the Poisson distribution with mean 70 are scattered independently over the square, each according to the uniform distribution. Note that the last plot from Figure 1.1 gives a visual impression rather similar to these six realizations.

That the Poisson process comparatively often gives a rather realistic model for observed data is reflected in the fact that there are many theorems that state convergence in distribution of a certain sequence of point processes to a Poisson process. Typically, the elements of these sequences are the result of modifying a certain point process (or several that have similar characteristics) according to a procedure that is more and more *entropy increasing* as the indices get higher and higher. This means, in a loose sense, that the amount of "randomness" in the original point process (in other words, the "additional information" contained in a realization as compared to the information already contained in the distribution of the process) is gradually increased, so that the point process tends to a Poisson process, which in a sense is the "most random" process possible. This idea can be made rigorous, at least in part, by the formal definition of point process entropy given

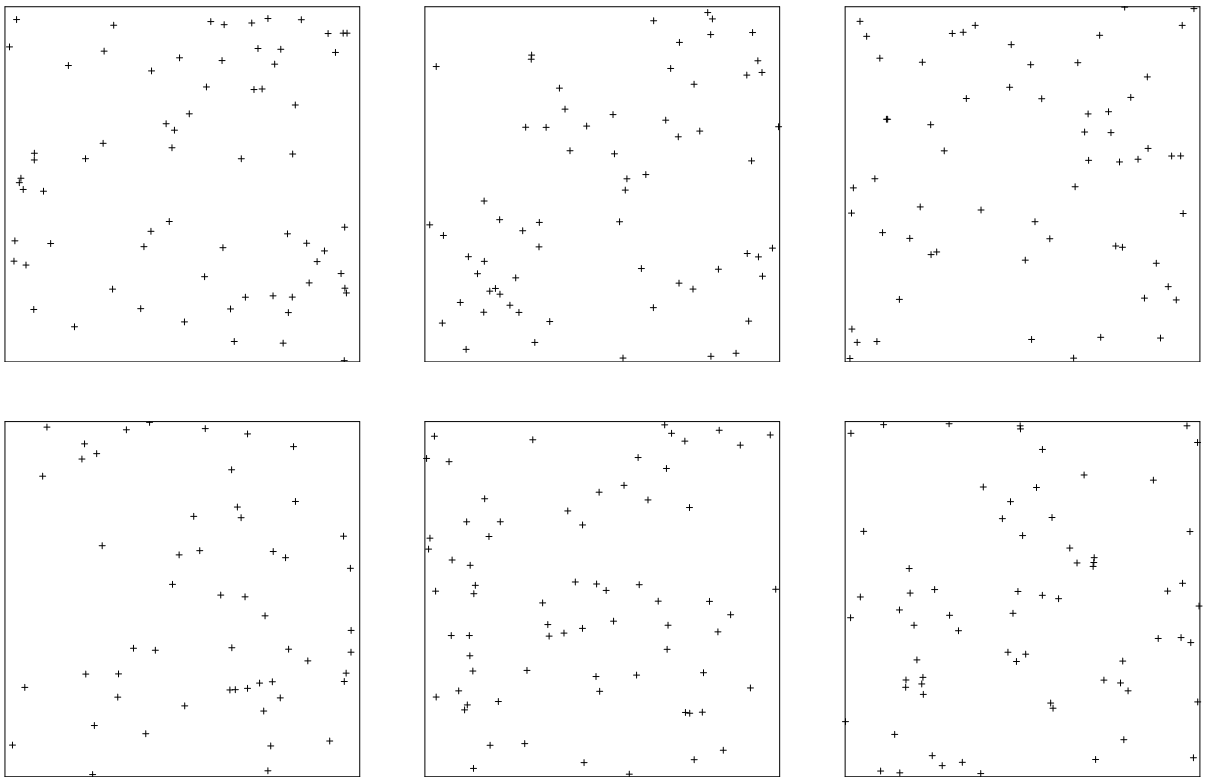

FIGURE 1.2: SIX DIFFERENT REALIZATIONS OF A HOMOGENEOUS POISSON PROCESS WITH INTENSITY 70 ON THE UNIT SQUARE.

for example in Daley and Vere-Jones (1988), Equation (13.5.7). Note also Exercise 13.5.2, which shows in an important special case that the Poisson process is the point process with the highest entropy among all processes with a given expectation measure (i.e. with the expected numbers of points in each measurable subset of the state space given).

In this thesis, we consider three different types of point process modifications that are entropy increasing, at least in an intuitive sense. Note that the following descriptions are rather imprecise and should be taken only as guidelines.

The first type of modification is formed by linear transformations on $\mathbb{R}^D$ that stretch some of the coordinate axes, while contracting the remainder. They are entropy increasing when applied to point processes that do not exhibit too much point clustering in the stretching directions and that show decreasing dependence between point positions with increasing distance in the contracting directions. The reason is essentially that weakly dependent parts of the point process are drawn closer from the contracting directions and strongly dependent parts are moved away in the stretching directions.

The second type of modification consists of thinnings that are combined with a compensating contraction of space. Thinnings are essentially defined as random deletions in a point process according to some rules. These modifications are entropy increasing when applied to processes that show decreasing dependence between point positions with increasing distance in any direction. The reason is similar to the one above: weakly dependent parts of the point process are drawn closer, while this time the thinning rule sees

to it that local dependences are suitably reduced.

The third type, finally, is given by superpositions of point processes in combination with a mechanism that makes the involved processes sparser. We have an increase of entropy compared to a "typical" single process if the latter does not exhibit too much dependence in the way points are retained under the sparseness mechanism, depends only weakly on most of the other processes and not all too strongly on the few remaining ones. Again the reason for the increased entropy is much as before: mostly weakly dependent point processes are superimposed (which is an extreme form of bringing them closer together), whereas the sparseness mechanism sees to it that the dependence within strongly dependent groups of processes is suitably reduced.

For all of the modifications described above, there are theorems readily available that give concrete conditions for convergence in distribution to a Poisson process, such as Corollary 2.7 in Ellis (1986) for the linear transforms, Theorem 2 in Brown (1979) for the thinnings (the setting there and also the limiting distribution are quite a bit more general than the ones described above), or Theorem 4 in Banys (1980) for the superposition.

However, a convergence theorem is rarely the end of the story. It is a very nice result for theoretical purposes, because it provides some insight into the fundamental relationships between the mathematical objects involved; but it fails to yield any quantifications of these relationships and thus in itself is close to useless for most practical applications.

In order to make the reasoning here a bit more concrete, consider a sequence (x_n) on a metric space (Γ, d), and assume that $x_n \to x$ for some element $x \in \Gamma$. Thus, in a somewhat loose sense, x_n gets arbitrarily close to x in terms of d if only we choose n large enough. But how close is it for any fixed n? The answer to this question is invaluable for practical computations, as well as when a certain theoretical property for a fixed x_n is to be proved by deriving it from the same property for x. However, a mere convergence result cannot provide this answer. It might just as well be that x_n is bounded away from x by a huge margin for every n that is reasonable in view of the practical situation one is facing (think of n, for example, as the sample size in a costly statistical experiment, and of x_n as the null hypothesis distribution of an appropriate statistic). On the other hand, even if (x_n) does not converge, it might well be that $d(x_n, x)$ is so small for some or even all n as to be negligible for the purpose intended.

From these considerations it becomes clear that the information contained in the distances $d(x_n, x)$ or in good upper estimates for them is enormously more valuable than a mere convergence result. The latter might have a certain appeal, as it is usually nicer to formulate and easier to prove, but wherever possible it should always be our goal to go for the concrete distance information, in order to make our results applicable.

It is the main focus of this thesis to give such distance information in the context of the three approximation settings mentioned above. More precisely, we derive upper bounds for distances between the distribution of a fairly general point process that is subjected to one of the three modifications presented and a suitable Poisson process distribution. As the principal metric for measuring these distances, the Barbour-Brown metric d_2 is chosen; this is a special kind of Wasserstein metric that has proved its worth in numerous applications. A detailed discussion of the d_2-metric with many of its properties and applications is a second focal point of this thesis.

The main distance results are proved using Stein's method, a very versatile technique by which many amazingly good distance estimates have been obtained with respect to a number of probability metrics and a large variety of different target distributions.

The thesis is organized as follows. Chapter 2 gives a detailed introduction to the mathematical concepts needed that are not typically part of a two semester course in measure theoretically founded probability theory. Most prominently, the elements of point process theory are presented, along with some of the most basic concepts from stochastic geometry, and an introduction to the general forms of the probability metrics used in this thesis. Chapter 3 discusses the more specific metrics on the space $\mathfrak{N}'$ of finite point measures on a compact set, and on the space $\mathfrak{P}(\mathfrak{N}')$ of probability measures on $\mathfrak{N}'$. In particular, the Barbour-Brown metric d_2 is introduced and its properties along with various applications are given. In Chapter 4, an overview of Stein's method is presented and some of the resulting general theorems about upper bounds for distances, especially between point process distributions, are listed.

With the preparations from Chapters 2 to 4, we are ready to tackle the three announced distance estimation problems. Chapters 5, 6, and 7 deal with distance estimates for the linear stretch-contract transforms, the thinnings and the superpositions, respectively. Each of these chapters is basically subdivided into three parts: first the problem is formulated, then various results about upper bounds are stated and proved, and finally a number of applications is given.

New results by the author are contained in Chapters 3, 5, 6, and 7. Chapters 2 and 4 have essentially been compiled from the literature about the various topics with only slight modifications and some additions every now and then. Chapter 3 contains many results about distances in connection with point measures and point processes that do not seem to be published, or, in a very few cases, are published but without formal proofs. This holds true most notably for the list of Lipschitz continuous functions in Section 3.3. Chapters 5 to 7 have been published by the author in the course of his PhD studies as Schuhmacher (2005a), (2005b), and (2005c). The contents of these chapters are in essence identical to the contents of the corresponding papers, though certain adaptations have been made to the introductory parts of the articles, and some notation has been changed for consistency reasons. Furthermore, Subsection 5.3.1 is new and presents a rather interesting connection between the three approximation problems.

Chapter 2

Probability prerequisites

In this chapter, we give the general theory needed for the main part as far as it lies beyond a "standard" two semester course in measure theoretically founded probability. First, some basic notation is introduced which holds for the entire thesis.

2.1 General notation and conventions

We start with the standard sets of numbers. Let $\mathbb{N} := \{1, 2, \ldots\}$, $\mathbb{Z}_+ := \{0, 1, 2, \ldots\}$, $\mathbb{R}_+ := [0, \infty)$ and define the sets $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ as usual. For any such set, a bar denotes that we include the symbols $+\infty$ and/or $-\infty$ as appropriate. Thus, for example, $\overline{\mathbb{Z}}_+ := \mathbb{Z}_+ \cup \{\infty\}$ or $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$. Where not stated otherwise, we make the assumptions that $\inf \emptyset = \infty$ and $\sup \emptyset = -\infty$, and that we use the canonical calculation rules with ∞ and $-\infty$.

We always denote by $(\Omega, \mathcal{A}, \mathbb{P})$ the underlying probability space on which our random elements are defined, and make the usual assumption that it is rich enough for all the objects to exist.

Let $(E, \mathcal{T})$ be a locally compact, second countable Hausdorff space (lcscH). This will be the most general state space that we consider for our point processes. Any such space is Polish (see Proposition A.2.A in the appendix), which means that there is a metric d inducing the topology $\mathcal{T}$ such that (E, d) is a complete, separable metric space. When dealing with metrics between point measures and point process distributions, we often require a compact metric space. We denote this space by (E', d_0), and require that $d_0 \leq 1$. Such a space is typically obtained by taking a compact subset of E together with the metric $\min(d|_{E'}, 1)$. Note that trimming of the metric has no influence on the topology induced. We will use various choices for the spaces E and E' throughout this thesis, but most frequently we set $E = \mathbb{R}^D$ or $E = \mathbb{R}^D_+$, endowed with the usual Euclidean topology, and define E' to be a closed hypercube in $\mathbb{R}^D$.

We consider the Borel σ-algebra on E and denote it by $\mathcal{B}$. For any subset $A \subset E$, denote by $\mathcal{B}_A$ the restriction of $\mathcal{B}$ to A, which is the Borel σ-algebra of the topological subspace A with the induced topology. If λ is a measure on E, $\mathcal{B}_\lambda$ denotes the algebra of λ-continuity sets, that is $\mathcal{B}_\lambda = \{A \in \mathcal{B};\, \lambda(\partial A) = 0\}$, where ∂A is the topological boundary of A. In the analogous way, we define $\mathcal{B}'$, $\mathcal{B}'_A$, and $\mathcal{B}'_\lambda$, based on the space E'. We call a set $B \subset E$ *(topologically) bounded* if it is relatively compact (i.e. its closure is compact). Note that in general metric spaces, this property is (usually strictly) stronger

than the property "bounded in the metric". However in finite dimensional vector spaces the two properties coincide.

The following notation is used for balls in E. For $a \in E$ and $r \geq 0$, denote by $\mathbb{B}(a,r) := \{s \in E;\, d(s,a) \leq r\}$ the closed d-ball with center in a and radius r. Sometimes we need the punctured ball $\dot{\mathbb{B}}(a,r) := \mathbb{B}(a,r) \setminus \{a\}$, and very rarely the open ball $\mathring{\mathbb{B}}(a,r)$ is used. In $E = \mathbb{R}^D$ we denote by

$$\alpha_D = \frac{\pi^{D/2}}{\Gamma(D/2+1)} \text{ and } \omega_D = \frac{2\pi^{D/2}}{\Gamma(D/2)}$$

the volume and the surface area of the Euclidean unit ball $\mathbb{B}(0,1)$, respectively.

We denote by $\mathfrak{F}_m$ the set of all measurable functions $f : E \to \mathbb{R}_+$, and by $\mathfrak{F}_{cc}$ the subset of all continuous functions with compact support. For any measure μ on E (or E'), we denote by $|\mu| := \mu(E) \in \overline{\mathbb{R}}_+$ (resp. $|\mu| := \mu(E') \in \overline{\mathbb{R}}_+$) its total mass. For any $f \in \mathfrak{F}_m$ or integrable $f : E \to \mathbb{R}$, we sometimes write $\mu(f)$ for the integral $\int_E f\, d\mu$. If X is a random element of some measurable space, we write $\mathscr{L}(X) := \mathbb{P}X^{-1}$ for its distribution, and if Y is another random element of the same space, we write $X \overset{\mathscr{D}}{=} Y$ if $\mathscr{L}(X) = \mathscr{L}(Y)$.

For subsets $A \subset \mathbb{R}^D$, $|A| := Leb^D(A)$ denotes the D-dimensional Lebesgue measure of A. For any finite set I, $|I| := \#I$ denotes the cardinality of I. The modulus notation is only used where this causes no confusion.

We make the convention that we may leave away the addition "almost surely" for equations and inequalities between random elements in situations where it is evident (and of no importance) that the corresponding relation does not hold pointwise. Most notably, we do this in situation where at least one of the random elements is defined only almost surely; for example, we might write that

$$\mathbb{E}(X \,|\, Y) = Z$$

for certain random elements X, Y, and Z.

Notation of the form $\sum^{\neq}_{i_1,\ldots,i_k \in I}$ or $\bigcup^{\neq}_{i_1,\ldots,i_k \in I}$ indicates that we consider only tuples of indices $(i_1,\ldots,i_k) \in I^k$ that are pairwise different (but in all $k!$ orders). Writing $\sum^{<}_{i_1,\ldots,i_k \in I}$ or $\bigcup^{<}_{i_1,\ldots,i_k \in I}$ means that we consider only pairwise different and *ordered* tuples, that is $(i_1,\ldots,i_k) \in I^k$ with $i_1 < \ldots < i_k$. If there is any notation at the top of $\sum$ or $\bigcup$, we separate the relation sign by a comma, as in $\sum^{v,<}_{i_1,\ldots,i_k=1}$.

2.2 Point processes

Point processes are the mathematical objects that are in the center of our attention for this thesis. In what follows, we give an overview of the results that are most important for our purposes. A more detailed account of point processes can be found in Kallenberg (1986).

2.2.1 Basic definitions

Intuitively, a point process is a random pattern of points on the space E, with the constraint that it may have only finitely many points in every (topologically) bounded set.

Formally, it is convenient to define a point process as a random locally finite point measure. A measure μ on E is called *locally finite* if $\mu(B) < \infty$ for every bounded $B \in \mathcal{B}$. A *point measure* is a $\overline{\mathbb{Z}}_+$-valued measure.

Denote by $\mathfrak{M}$ the set of all locally finite measures on E, and by $\mathfrak{N}$ the set of all locally finite point measures on E. We endow these spaces with their vague topologies $\mathcal{T}_{\mathfrak{M}}$ and $\mathcal{T}_{\mathfrak{N}}$, respectively (see Subsection A.2.2 in the Appendix for the definitions and for the next few results). Furthermore, we define the σ-algebras $\mathcal{M}$ and $\mathcal{N}$ as the smallest ones that render the mass functions of bounded measurable sets measurable, that is we set

$$\mathcal{M} := \sigma\big([\mathfrak{M} \ni \mu \mapsto \mu(B)];\ B \in \mathcal{B} \text{ bounded}\big) \text{ and}$$
$$\mathcal{N} := \sigma\big([\mathfrak{N} \ni \mu \mapsto \mu(B)];\ B \in \mathcal{B} \text{ bounded}\big).$$

We can now define point processes or, more generally, random measures as follows.

Definition.

(a) A measurable mapping $\Lambda : \Omega \to \mathfrak{M}$ is called a *random measure* on E.

(b) A measurable mapping $\xi : \Omega \to \mathfrak{N}$ is called a *point process* on E.

The above choices of σ-algebras are natural in two senses. First, they guarantee that if we have a random measure Λ or a point process ξ, then the masses $\Lambda(B)$ and the point counts $\xi(B)$ are random variables for every bounded $B \in \mathcal{B}$ (suitably formulated conversions hold as well; see Daley and Vere-Jones (1988), Propositions 6.1.III and 7.1.VIII). Secondly, they go well with the topologies introduced, inasmuch as $\mathcal{M}$ and $\mathcal{N}$ are the Borel-σ-algebras of $(\mathfrak{M}, \mathcal{T}_{\mathfrak{M}})$ and $(\mathfrak{N}, \mathcal{T}_{\mathfrak{N}})$, respectively.

In this thesis, the topic of general random measures is touched only briefly. We therefore concentrate on point processes in what follows. Sometimes it is easier to deal with point processes that have at most one point at every position of the space E.

Definition. A point process ξ is called *simple* if $\mathbb{P}\big[\xi(\{s\}) \geq 2 \text{ for some } s \in E\big] = 0$.

The set $\big\{\varrho \in \mathfrak{N}; \varrho(\{s\}) \geq 2 \text{ for some } s \in E\big\}$ can be shown to be in $\mathcal{N}$, because of the separability of E.

Let now $E := \mathbb{R}^D$. Stationarity of point processes is a useful property, since it simplifies certain theoretical considerations and permits statistical inference for point process characteristics based on a single realization. Set $s + A := \{s + a; a \in A\}$ for any $s \in \mathbb{R}^D$ and any $A \subset \mathbb{R}^D$. We consider only the strong form of stationarity, which is given as follows.

Definition. A point process ξ is called *stationary* if for any bounded sets $B_1, \ldots, B_r \in \mathcal{B}$, $r \in \mathbb{N}$, and for any $s \in \mathbb{R}^D$,

$$\mathscr{L}\big(\xi(s + B_1), \ldots, \xi(s + B_r)\big) = \mathscr{L}\big(\xi(B_1), \ldots, \xi(B_r)\big).$$

2.2.2 Finite dimensional distributions

Instead of dealing with the distribution of a point process, which is a rather complex object (a probability measure on $\mathfrak{N}$), it is often easier to consider the so-called finite dimensional distributions, which are just probability measures on $\mathbb{Z}_+^r$.

Definition. Let ξ be a point process on E. Then for any bounded sets $B_1, \dots B_r \in \mathcal{B}$, $r \in \mathbb{N}$, we call

$$\mathscr{L}\big(\xi(B_1), \dots, \xi(B_r)\big)$$

an *r-th finite dimensional distribution* (or short *fidi-distribution*) of ξ.

The following proposition is helpful in order to prove that a point process has a certain distribution.

Proposition 2.2.A. *Let ξ_1 and ξ_2 be point processes on E which have the same fidi-distributions for pairwise disjoint sets; that is, they satisfy*

$$\mathbb{P}\big[\xi_1(B_1) = k_1, \dots, \xi_1(B_r) = k_r\big] = \mathbb{P}\big[\xi_2(B_1) = k_1, \dots, \xi_2(B_r) = k_r\big]$$

for all $r \in \mathbb{N}$, all bounded and pairwise disjoint $B_1, \dots, B_r \in \mathcal{B}$, and all $k_1, \dots, k_r \in \mathbb{Z}_+$. Then $\xi_1 \stackrel{\mathscr{D}}{=} \xi_2$.

Proof. Let $\mathcal{C}$ be the system of all sets of the form

$$\big\{\varrho \in \mathfrak{N};\ \varrho(B_1) = k_1, \dots, \varrho(B_r) = k_r\big\}$$

with $r \in \mathbb{N}$, $B_1, \dots, B_r \in \mathcal{B}$ bounded, and $k_1, \dots, k_r \in \mathbb{Z}_+$. Since every $C \in \mathcal{C}$ can be written as a finite disjoint union of sets of the same form, but with $B_1, \dots, B_r \in \mathcal{B}$ bounded *and pairwise disjoint*, we obtain by the prerequisite that $\mathbb{P}[\xi_1 \in C] = \mathbb{P}[\xi_2 \in C]$ for every $C \in \mathcal{C}$. But $\mathcal{C}$ is obviously a π-system (i.e. closed with respect to intersections) that generates $\mathcal{N}$ (i.e. $\sigma(\mathcal{C}) = \mathcal{N}$), whence it follows by a standard result from measure theory (see e.g. Kallenberg (2002), Lemma 1.17) that $\mathbb{P}\xi_1^{-1} = \mathbb{P}\xi_2^{-1}$. □

2.2.3 Convergence in distribution

Of course, convergence in distribution of point processes is only a special case of convergence in distribution for more general random elements. We give the explicit definition here for the sake of completeness

Definition. Let $\xi, \xi_1, \xi_2, \dots$ be point processes on E. Then we say that ξ_n *converges in distribution* to ξ (or more precisely that $\mathscr{L}(\xi_n)$ *converges weakly* to $\mathscr{L}(\xi)$) and write

$$\xi_n \stackrel{\mathscr{D}}{\longrightarrow} \xi \quad (n \to \infty)$$

if

$$\mathbb{E}f(\xi_n) \longrightarrow \mathbb{E}f(\xi) \quad (n \to \infty)$$

for every bounded continuous function $f : \mathfrak{N} \to \mathbb{R}$, where continuity is to be understood with respect to our usual topology $\mathcal{T}_{\mathfrak{N}}$ on $\mathfrak{N}$.

There are several criteria for the convergence in distribution of a sequence of point processes. We present here only the convergence of the fidi-distributions.

Proposition 2.2.B. *Let $\xi, \xi_1, \xi_2, \dots$ be point processes on E. We have $\xi_n \stackrel{\mathscr{D}}{\longrightarrow} \xi$ if and only if*

$$\big(\xi_n(B_1), \dots, \xi_n(B_r)\big) \stackrel{\mathscr{D}}{\longrightarrow} \big(\xi(B_1), \dots, \xi(B_r)\big)$$

for all $r \in \mathbb{N}$ and all bounded $B_1, \dots, B_r \in \mathcal{B}$ that are stochastic continuity sets with respect to ξ (that is, the B_i satisfy $\xi(\partial B_i) = 0$ a.s.).

Proof. The statement follows from Theorem 4.2 and Lemma 4.3 in Kallenberg (1986). □

2.2.4 Representations

Our intuitive idea behind a point process is that of a random number of points scattered randomly over the set E. It is often desirable to formulate events based on these randomly scattered points; that is, we would like to have random elements that describe the positions of the points in a given point process. That this is possible is guaranteed by the next two results.

Proposition 2.2.C. *Every point process ξ on E may be written in the form*

$$\xi = \sum_{i=1}^{V} X_i \delta_{S_i}$$

with $\sigma(\xi)$-measurable random elements V with values in $\overline{\mathbb{Z}}_+$, $X_1, X_2, \ldots$ with values in $\mathbb{N}$, and $S_1, S_2, \ldots$ with values in E. The representation may be chosen in such a way that $S_1(\omega), S_2(\omega), \ldots$ are pairwise different for every $\omega \in \Omega$. In addition, for an arbitrary partition $(B_j)_{j\in\mathbb{N}}$ of E into bounded measurable sets, the representation may be chosen B_j-wise, meaning that it may be chosen in such a way that $S_1(\omega), \ldots, S_{\xi(B_1)(\omega)}(\omega) \in B_1$, $S_{\xi(B_1)(\omega)+1}(\omega), \ldots, S_{\xi(B_1)(\omega)+\xi(B_2)(\omega)}(\omega) \in B_2$, and so on.

Proof. The representation is an immediate consequence of Lemmas 2.1 and 2.2 in Kallenberg (1986). That the special choices for the representations are possible follows directly from the construction in the proof of Lemma 2.2. □

Corollary 2.2.D. *Every point process ξ on E may be written in the form*

$$\xi = \sum_{i=1}^{V} \delta_{S_i},$$

where V and $S_1, S_2, \ldots$ are $\sigma(\xi)$-measurable random elements with values in $\overline{\mathbb{Z}}_+$ and E, respectively. For an arbitrary partition $(B_j)_{j\in\mathbb{N}}$ of E into bounded measurable sets, the representation may be chosen B_j-wise as in Proposition 2.2.C.

Proof. Choose a B_j-wise representation $\sum_{i=1}^{\widetilde{V}} X_i \delta_{\widetilde{S}_i}$ for ξ as in Proposition 2.2.C. Set

$$K_l(\omega) := \min\left\{ k \in \mathbb{N};\ \sum_{i=1}^{k} X_i(\omega) \geq l \right\}$$

and $S_l(\omega) := \widetilde{S}_{K_l(\omega)}(\omega)$ for every $l \in \mathbb{N}$. Write furthermore $V(\omega) := \sum_{i=1}^{\widetilde{V}(\omega)} X_i(\omega)$. Then V and $S_1, S_2, \ldots$ are $\sigma(\xi)$-measurable random elements of E, and

$$\sum_{i=1}^{V} \delta_{S_i} = \xi$$

is a B_j-wise representation of ξ. □

Whenever we write from now on a point process as $\xi = \sum_{i=1}^{V} \delta_{S_i}$, it is tacitly assumed that V and $S_1, S_2, \ldots$ are $\sigma(\xi)$-measurable random elements of $\overline{\mathbb{Z}}_+$ and E, respectively (unless we explicitly specify $V, S_1, S_2, \ldots$ otherwise).

2.2.5 Moment measures

For real-valued random variables we have the first few moments (or the cumulants, like e.g. the variance) as important "summary numbers" for the information contained in their distributions. For point processes, similar summaries can be defined, but in this case the summaries are measures on product spaces of E.

Definition. Let ξ be a point process on E and $k \in \mathbb{N}$. Then we call the measure μ_k on $(E^k, \mathcal{B}^k)$ that is defined on the rectangles by

$$\mu_k(A_1 \times \ldots \times A_k) := \mathbb{E}\big(\xi(A_1) \cdots \xi(A_k)\big)$$

for $A_1, \ldots, A_k \in \mathcal{B}$, the *k-th moment measure* of ξ. In particular, we call μ_1 the expectation measure of ξ and denote it by $\mathbb{E}\xi$.
We say that the k-th moment measure *exists* if it is locally finite.

Following the real-valued case, one sometimes considers the *covariance measure* of a point process, which, for finite second moment measure, is given by $\mathrm{cov}(\xi) := \mu_2 - \mu_1^2$, and hence is a signed measure on $(E^2, \mathcal{B}^2)$.

The k-th moment measure can obviously be interpreted as the expectation measure of the product point process ξ^k on E^k. This makes the major shortcoming of the moment measure μ_k for $k \geq 2$ evident: except for the trivial case that $\xi = 0$ almost surely, μ_k always has positive mass on the diagonal $\{(s_1, \ldots, s_k) \in E^k;\, s_1 = \ldots = s_k\}$ (in fact, on any "diagonal subspace" $\{(s_1, \ldots, s_k) \in E^k;\, s_{i_1} = \ldots = s_{i_r}\}$ of E, where $i_1, \ldots, i_r \in \{1, 2, \ldots, k\}$ are pairwise different with $r \geq 2$), which is an unattractive degeneracy. Suppose that $E = \mathbb{R}^D$. Then it may well be that the expectation measure of ξ is absolutely continuous with respect to Leb^D, but any higher moment measure μ_k is never absolutely continuous with respect to Leb^{Dk} (except in the trivial case where $\xi = 0$). This problem can be remedied by considering factorial moment measures, where the problematic diagonal masses have been removed.

Definition. Let ξ be a point process on E and $k \in \mathbb{N}$. For $m \in \mathbb{Z}_+$, write $m^{[k]} := m(m-1)\ldots(m-k+1)$. Then we call the measure $\mu_{[k]}$ on $(E^k, \mathcal{B}^k)$ that is defined on the rectangles via

$$\mu_{[k]}(A_1^{k_1} \times \ldots \times A_r^{k_r}) := \mathbb{E}\big(\xi(A_1)^{[k_1]} \cdots \xi(A_r)^{[k_r]}\big),$$

where $r \in \mathbb{N}$, $k_1, \ldots, k_r \in \mathbb{N}$ with $\sum_{i=1}^r k_i = k$, and $A_1, \ldots, A_r \in \mathcal{B}$ pairwise disjoint, the *k-th factorial moment measure* of ξ.

It is easily seen that the second factorial moment measure is the expectation measure of

$$\xi^2 - \sum_{i=1}^{V} \delta_{(S_i, S_i)},$$

and, more generally, that the k-th factorial moment measure $\mu_{[k]}$ is the expectation measure of

$$\sum_{i_1, \ldots, i_k = 1}^{V, \neq} \delta_{(S_{i_1}, \ldots, S_{i_k})},$$

where $\sum_{i=1}^{V} \delta_{S_i}$ is an arbitrary representation of ξ.

We finish this subsection by giving different versions of a useful formula that involves moment measures.

Proposition 2.2.E (Campbell's Formula). *Let ξ be a point process on E and $f \in \mathfrak{F}_m$. Then*

$$\mathbb{E}\left(\int f(s)\ \xi(ds)\right) = \int f(s)\ \mu_1(ds)\ \in \overline{\mathbb{R}}_+.$$

Proof. The proof follows from a simple extension argument. We give it here once in all detail. Later on, we often leave the particulars of similar extension arguments for the reader to check. First, let $f := 1_A$ be an indicator function with $A \in \mathcal{B}$. Then the statement is just the definition of the expectation measure. The statement holds for simple functions, that is functions of the form $\sum_{i=1}^{n} a_i 1_{A_i}$ with $a_i \geq 0$, $A_i \in \mathcal{B}$, by the linearity of integral and expectation. We then use the fact that if f is a non-negative measurable function it can be approximated from below by non-negative simple functions (see e.g. Kallenberg (2002), Lemma 1.11). Hence for a sequence $(f_n)_{n\in\mathbb{N}}$ of simple functions with $f_n \nearrow f \in \mathfrak{F}_m$,

$$\mathbb{E}\left(\int f(s)\ \xi(ds)\right) = \lim_{n\to\infty} \mathbb{E}\left(\int f_n(s)\ \xi(ds)\right) = \lim_{n\to\infty} \int f_n(s)\ \mu_1(ds) = \int f(s)\ \mu_1(ds).$$

by using Levi's monotone convergence theorem several times. □

Corollary 2.2.F. *Let ξ be a point process on E and $f \in \mathfrak{F}_m$.*

(i) For any $k \in \mathbb{N}$ we have

$$\mathbb{E}\left(\int f(s)\ \xi(ds)\right)^k = \int f(s_1)\cdots f(s_k)\ \mu_k(ds_1\ \ldots\ ds_k)\ \in \overline{\mathbb{R}}_+.$$

(ii) If the second moment measure μ_2 is finite, we have

$$\mathrm{var}\left(\int f(s)\ \xi(ds)\right) = \int f(s_1)f(s_2)\ \mathrm{cov}(\xi)(ds_1\, ds_2)\ \in \overline{\mathbb{R}}_+.$$

Proof. (i) Apply Proposition 2.2.E to

$$\left(\int_E f(s)\ \xi(ds)\right)^k = \int_{E^k} f(s_1)\cdots f(s_k)\ \xi^k(ds_1\ \cdots\ ds_k).$$

(ii) By (i), we have

$$\begin{aligned}
\mathrm{var}\left(\int f(s)\ \xi(ds)\right) &= \mathbb{E}\left(\left(\int f(s)\ \xi(ds)\right)^2\right) - \left(\mathbb{E}\left(\int f(s)\ \xi(ds)\right)\right)^2 \\
&= \int f(s_1)f(s_2)\ \mu_2(ds_1\, ds_2) - \int f(s_1)f(s_2)\ \mu_1^2(ds_1\, ds_2) \\
&= \int f(s_1)f(s_2)\ \mathrm{cov}(\xi)(ds_1\, ds_2).
\end{aligned}$$

□

2.2.6 Generating functionals

Generating functionals are a powerful tool for characterizing point process distributions. They are related to generating functions of $\mathbb{Z}_+$-valued random variables, and can be regarded as somewhat more accessible versions of the so-called Laplace functionals for general random measures (which are the counterpart of the Laplace functions for real-valued random variables).

The strength of generating functionals lies in their relative simplicity. For showing distributional properties of a point process, it is often easier to do calculations with the generating functionals than with the complicated point process distributions directly. The downside of the generating functional approach lies in the fact that it is often not very easy to grasp intuitively, so that proofs done with the aid of generating functionals are usually just analytical calculations, from which little insight is gained into the mechanisms behind the result.

For this reason, we mostly avoid using generating functionals, although some of the proofs in this chapter could be carried out more quickly with their aid. To illustrate this point, we give two proofs for Proposition 2.2.S below, one using a more straightforward idea, and one using generating functionals. Note that the proof using generating functionals is quite a bit more elegant, but the meaning of the different steps is rather obscured.

Let $\mathfrak{F}_u := \big\{h \in \mathfrak{F}_m;\, 0 \le h \le 1, \{s \in E; h(s) < 1\} \text{ is bounded}\big\}$.

Definition. For any point process $\xi = \sum_{i=1}^{V} \delta_{S_i}$ on E, we define its *generating functional* $G : \mathfrak{F}_u \to [0,1]$ by

$$G(h) := G_\xi(h) := \mathbb{E}\bigg[\exp\Big(\int_E \log\big(h(s)\big)\, \xi(ds)\Big)\bigg] = \mathbb{E}\Big(\prod_{i=1}^{V} h(S_i)\Big)$$

for every $h \in \mathfrak{F}_u$

The last equality is obtained by a very simple calculation. Note that, since $h \in \mathfrak{F}_u$, the integral above can be written as a finite (random) sum, and the product consists of only a finite (random) number of factors that are different from 1.

As mentioned above, it can be shown that the distribution of a point process is characterized by its generating functional.

Proposition 2.2.G. *Let ξ_1 and ξ_2 be point processes on E which both have the same generating functional G. Then $\xi_1 \stackrel{\mathcal{D}}{=} \xi_2$.*

The proof of this result, as well as more results about generating functionals, can be found in Daley and Vere-Jones (1988), Section 7.4. We finish this subsection by showing that the generating functional is a generalization of the generating function of a $\mathbb{Z}_+$-valued random variable.

Proposition 2.2.H. *Let $\xi = \sum_{i=1}^{V} \delta_{S_i}$ be a point process on E, and define $h \in \mathfrak{F}_u$ by $h(s) := 1 - \sum_{j=1}^{m}(1 - z_j)1_{B_j}(s)$ for $m \in \mathbb{N}$, $z_j \in [0,1]$, and for $B_j \in \mathcal{B}$ bounded and pairwise disjoint. Then*

$$G_\xi(h) = \mathbb{E}\Big(\prod_{j=1}^{m} z_j^{\xi(B_j)}\Big),$$

which as a function of $(z_1, \dots, z_m)$ *is just the multivariate generating function of the random vector* $(\xi(B_1), \dots, \xi(B_m))$. *In particular, we obtain for* $m = 1$ *the generating function of* $\xi(B_1)$.

Proof. For any $s \in E$, set $r(s) := 0$ if $s \notin \bigcup_{j=1}^m B_j$ and define $r(s) \in \{1, 2, \dots, m\}$ by $s \in B_{r(s)}$ otherwise. Set furthermore $z_0 := 1$. Then

$$G_\xi(h) = \mathbb{E}\Big(\prod_{i=1}^V \Big(\sum_{j=1}^m z_j 1_{B_j}(S_i)\Big)\Big) = \mathbb{E}\Big(\prod_{i=1}^V z_{r(S_i)}\Big) = \mathbb{E}\Big(\prod_{j=1}^m z_j^{\xi(B_j)}\Big).$$

□

Thus the generating functional of a point process can be interpreted as a limiting version of the multivariate generating function of the vector of point counts, when the space E is subdivided into infinitesimal regions.

2.2.7 Binomial, Poisson and Cox processes

In this subsection, we present some important types of point processes. We first restrict ourselves to finite versions of these processes. Remember that we use $|\lambda| := \lambda(E)$ to denote the total mass of a measure λ on E.

Definition. Let ν and λ be finite measures on E, where $|\nu| = 1$.

(a) Let $n \in \mathbb{Z}_+$, and let S_i be independent E-valued random elements with distribution ν. We then call $\eta := \sum_{i=1}^n \delta_{S_i}$ a *Binomial process* on E with size n and point distribution ν, and write $\eta \sim \mathrm{Bi}(n, \nu)$.

(b) We call η a *finite Poisson process* on E with intensity measure λ and write $\eta \sim \mathrm{Po}(\lambda)$ if $\eta(E)$ is Poisson distributed with parameter $|\lambda| \geq 0$, and if, given $\eta(E) = n$, η is a $\mathrm{Bi}(n, \lambda/|\lambda|)$-process for $|\lambda| > 0$ (note that $\eta(E) = 0$ a.s. if $|\lambda| = 0$).

Since the definitions imply explicit constructions of the point processes, it follows immediately that they exist. Note that we use the notations $\mathrm{Bi}(n, \cdot)$ and $\mathrm{Po}(\cdot)$ to denote either point process distributions or distributions of $\mathbb{Z}_+$-valued random variables ("point processes on a singleton"), depending on whether the parameter is a measure or a real number.

Next, we give the fidi-distributions of the two processes for pairwise disjoint sets. Note that both processes are named after their one-dimensional distributions.

Proposition 2.2.I. *The fidi-distributions of the* $\mathrm{Bi}(n, \nu)$*-process* η *on* E *are given by*

$$\mathbb{P}\big[\eta(B_1) = k_1, \dots, \eta(B_r) = k_r\big] = \begin{cases} \frac{n!}{k_1! \cdots k_r!} \nu(B_1)^{k_1} \cdots \nu(B_r)^{k_r} & \text{if } \sum_{i=1}^r k_i = n, \\ 0 & \text{otherwise,} \end{cases}$$

where $r \in \mathbb{N}$, $(B_i)_{1 \leq i \leq r}$ *is a finite partition of* E *into measurable sets, and* $k_1, \dots, k_r \in \mathbb{Z}_+$. *Hence* $\big(\eta(B_1), \dots, \eta(B_r)\big)$ *is multinomially distributed with size parameter* n *and probability vector* $(\nu(B_1), \dots, \nu(B_r))$.

Proof. This is a direct consequence of the definition. □

Proposition 2.2.J. *The fidi-distributions of the* $\mathrm{Po}(\lambda)$*-process* η *on* E *are given by*

$$\mathbb{P}\big[\eta(B_1) = k_1, \dots, \eta(B_r) = k_r\big] = \frac{\lambda(B_1)^{k_1}}{k_1} e^{-\lambda(B_1)} \dots \frac{\lambda(B_r)^{k_r}}{k_r} e^{-\lambda(B_r)},$$

where $r \in \mathbb{N}$, $B_1, \dots, B_r \in \mathcal{B}$ *are disjoint sets, and* $k_1, \dots, k_r \in \mathbb{Z}_+$.
Hence $\eta(B_1), \dots, \eta(B_r)$ *are independent with* $\eta(B_i) \sim \mathrm{Po}(\lambda(B_i))$.

Proof. Let $r \in \mathbb{N}$, $B_1, \dots, B_r \in \mathcal{B}$ disjoint, and $k_1, \dots, k_r \in \mathbb{Z}_+$. Furthermore, let $B := \bigcup_{i=1}^r B_i$, and $k := \sum_{i=1}^r k_i$. Then we have

$$\begin{aligned}
&\mathbb{P}[\eta(B_1) = k_1, \dots, \eta(B_r) = k_r] \\
&\quad = \sum_{n=0}^{\infty} \mathbb{P}\big[\eta(B_1) = k_1, \dots, \eta(B_r) = k_r \,\big|\, \eta(E) = n\big]\, \mathbb{P}[\eta(E) = n] \\
&\quad = \sum_{n=k}^{\infty} \frac{n!}{k_1! \cdots k_r! (n-k)!} \left(\frac{\lambda(B_1)}{|\lambda|}\right)^{k_1} \cdots \left(\frac{\lambda(B_r)}{|\lambda|}\right)^{k_r} \left(\frac{\lambda(B^c)}{|\lambda|}\right)^{n-k} \frac{|\lambda|^n}{n!} e^{-|\lambda|} \\
&\quad = \frac{\lambda(B_1)^{k_1}}{k_1!} e^{-\lambda(B_1)} \dots \frac{\lambda(B_r)^{k_r}}{k_r!} e^{-\lambda(B_r)} \sum_{n=k}^{\infty} \frac{\lambda(B^c)^{n-k}}{(n-k)!} e^{-\lambda(B^c)} \\
&\quad = \frac{\lambda(B_1)^{k_1}}{k_1!} e^{-\lambda(B_1)} \dots \frac{\lambda(B_r)^{k_r}}{k_r!} e^{-\lambda(B_r)},
\end{aligned}$$

where the second equality is obtained by Proposition 2.2.I. □

The property that point counts of disjoint sets are independent makes the Poisson process particularly nice to deal with. It enables us to obtain many simple and beautiful mathematical results (see further below in this subsection). But the Poisson process is not only important for its theoretical properties. Although the independence requirement is usually too strong to be completely satisfied in practical applications, it can be shown that in many situations, a Poisson process approximates a given more general process sufficiently well, which makes it a popular modeling tool.

We now extend the concept of a Poisson process to processes with general locally finite intensity measures. This is done most conveniently via the characterization in Proposition 2.2.J.

Definition (General Poisson process). Let η be a point process and $\lambda \in \mathfrak{M}$ a locally finite measure on E.
Then η is called a *Poisson process* with intensity measure λ if

(a) $\eta(B) \sim \mathrm{Po}(\lambda(B))$ for every bounded $B \in \mathcal{B}$;

(b) $\eta(B_1), \dots, \eta(B_r)$ are independent for every selection of bounded and pairwise disjoint sets $B_1, \dots, B_r \in \mathcal{B}$, $r \in \mathbb{N}$.

By Proposition 2.2.A, the above definition determines the distribution of the $\mathrm{Po}(\lambda)$-process uniquely. The existence of such a process is given by the explicit construction in the following result. Note that for compact E the construction includes the one from the definition of the finite Poisson process as a special case.

Proposition 2.2.K. *Let $\lambda \in \mathfrak{M}$, and let $(E_i)_{i\in\mathbb{N}}$ be an arbitrary partition of E into bounded measurable sets (such a partition exists by Proposition A.2.B), where $E_i = \emptyset$ is allowed for any of the sets. For each $i \in \mathbb{N}$, construct a finite $\mathrm{Po}(\lambda|_{E_i})$-process η_i on E_i in such a way that $(\eta_i)_{i\in\mathbb{N}}$ is an independent sequence. Then η, defined by $\eta(A) := \sum_{i\in\mathbb{N}} \eta_i(A \cap E_i)$ for every $A \in \mathcal{B}$, is a $\mathrm{Po}(\lambda)$-process on E.*

Proof. We show that the two properties from the above definition are satisfied.

(a) Let $B \in \mathcal{B}$ bounded. We have $\eta(B) = \sum_{i=1}^{\infty} \eta_i(B \cap E_i)$, where the $\eta_i(B \cap E_i)$ are independent (because the η_i are independent) and $\mathrm{Po}(\lambda(B\cap E_i))$-distributed (because η_i is a $\mathrm{Po}(\lambda|_{E_i})$-process). Hence $\eta(B)$ is Poisson-distributed with parameter $\sum_{i=1}^{\infty} \lambda(B\cap E_i) = \lambda(B)$ (for finitely many summands this is a well-known result; here we need an additional, but very simple, limit argument).

(b) Let $r \in \mathbb{N}$, and let $B_1, \dots, B_r \in \mathcal{B}$ be bounded disjoint sets. Then $\eta(B_j) = \sum_{i=1}^{\infty} \eta_i(B_j \cap E_i)$. Since η_i are Poisson processes that are independent of one another, it follows that $\eta_i(B_j \cap E_i)$, $1 \le j \le r$, $i \in \mathbb{N}$, are independent. Thus $\eta(B_j)$, $1 \le j \le r$, are independent, because they are functions of disjoint selections from $\eta_i(B_j \cap E_i)$, $1 \le j \le r$, $i \in \mathbb{N}$. □

Using the same proof idea for sets $A \in \mathcal{B}$ that are not necessarily bounded, yields

Proposition 2.2.L. *Let η be a Poisson process on E with intensity measure λ. Then*

(i) $\eta(A) \sim \mathrm{Po}(\lambda(A))$ for every $A \in \mathcal{B}$ with $\lambda(A) < \infty$, and $\eta(A) = \infty$ a.s. for every $A \in \mathcal{B}$ with $\lambda(A) = \infty$;

(ii) $\eta(A_1), \dots, \eta(A_r)$ are independent for pairwise disjoint sets $A_1, \dots, A_r \in \mathcal{B}$, $r \in \mathbb{N}$.

Proof. (i) Let $(B_i)_{i\in\mathbb{N}}$ be a partition of A into bounded Borel sets, so that $\eta(A) = \sum_{i=1}^{\infty} \eta(B_i)$. Since the $\eta(B_i)$ are independent and $\mathrm{Po}(\lambda(B_i))$-distributed, we have that $\sum_{i=1}^{n} \eta(B_i) \sim \mathrm{Po}\big(\sum_{i=1}^{n} \lambda(B_i)\big)$ for any $n \in \mathbb{N}$, whence we obtain statement (i) by a simple limit argument.

(ii) For every $j \in \{1, 2, \dots, r\}$, let $(B_{ji})_{i\in\mathbb{N}}$ be a partition of A_j into bounded Borel sets. Then B_{ji}, $1 \le j \le r$, $i \in \mathbb{N}$, are pairwise disjoint, whence we obtain the independence of $\eta(B_{ji})$, $1 \le j \le r$, $i \in \mathbb{N}$. Thus $\eta(A_j)$, $1 \le j \le r$, are independent, because they are functions of disjoint selections from $\eta(B_{ji})$, $1 \le j \le r$, $i \in \mathbb{N}$. □

By randomization, a much larger class of point processes based on the Poisson process can be defined.

Definition. Let Λ be a random measure on E. We call η a *Cox process* on E with directing measure Λ and write $\eta \sim \mathrm{Cox}(\Lambda)$ if $\mathscr{L}(\eta \,|\, \Lambda = \lambda) = \mathrm{Po}(\lambda)$ for Λ-a.e. λ.

As usually with mixed distributions, the uniqueness is immediately clear, and the existence follows in a straightforward manner e.g. from Kallenberg (1986), Theorem 15.3.2. Cox processes play a less important part in this thesis. We encounter them again in connection with random thinnings in Subsection 2.2.8 and Section 6.2.

* * *

To illustrate the general concepts for point processes introduced so far in this chapter, and to demonstrate the advantages of Poisson processes, we next calculate some of their

characteristic objects. In what follows, η is always a Poisson process on E with intensity measure $\lambda \in \mathfrak{M}$.

Proposition 2.2.M (Representation). *Let $B \in \mathcal{B}$ be bounded. Consider a random variable $V \sim \mathrm{Po}(\lambda(B))$ and i.i.d. random elements $S_1, S_2, \ldots$ with distribution $\lambda/\lambda(B)$ that are independent also of V. Then*

$$\sum_{i=1}^{V} \delta_{S_i} \stackrel{\mathscr{D}}{=} \eta|_B.$$

Proof. This is an immediate consequence of Proposition 2.2.K. □

Remark 2.2.N. The representation given in Proposition 2.2.M is the most general and convenient one for a Poisson process. However, it is for most choices of B and λ not a representation in the sense of Corollary 2.2.D, not even if all random variables are constructed on the same probability space in such a way that $\sum_{i=1}^{V} \delta_{S_i} = \eta|_B$. This is because the S_i cannot be written as functions of η, except in trivial cases.
For a homogeneous Poisson process η on $E = \mathbb{R}_+$ with intensity $\lambda > 0$ (that is, η is a $\mathrm{Po}(\lambda Leb|_{\mathbb{R}_+})$-process) a representation in the sense of Corollary 2.2.D is the classical representation $\eta = \sum_{i=1}^{\widetilde{V}} \delta_{\widetilde{S}_i}$, where $\widetilde{S}_i$ denotes the i-th point of η with respect to the natural order relation on $\mathbb{R}_+$. Note that $\widetilde{V} = \infty$ a.s., and $(\widetilde{S}_i)_i \stackrel{\mathscr{D}}{=} \left(\sum_{j=1}^{i} T_j\right)_i$ for i.i.d. T_j with $T_j \sim \mathrm{NE}(\lambda)$.
This representation can be extended to the homogeneous Poisson process η on $E = \mathbb{R}^D$ with intensity $\lambda > 0$ (i.e. $\eta \sim \mathrm{Po}(\lambda Leb^D)$). Write $\eta = \sum_{i=1}^{\widetilde{V}} \delta_{\widetilde{S}_i}$, where $\widetilde{S}_i$ denotes the point of η that is in the Euclidean metric i-th closest to the origin (with a reasonable secondary criterion for ties). Let U_i, $i \in \mathbb{N}$, be independent random vectors that are uniformly distributed on the unit sphere $\{s \in \mathbb{R}^D; |s| = 1\}$, and let T_j, $j \in \mathbb{N}$, be i.i.d. random variables with $T_j \sim \mathrm{NE}(\lambda\alpha_D)$ that are independent also of the U_i. Then $(\widetilde{S}_i)_i \stackrel{\mathscr{D}}{=} \left((\sum_{j=1}^{i} T_j)U_i\right)_i$. See Møller and Waagepetersen (2004), Section 3.1.3, for details.

Proposition 2.2.O (First and second moment measures). *Denote by λ_Δ the measure on E^2 which is defined by $\lambda_\Delta(A) = \lambda(\mathrm{pr}_1(A \cap \Delta))$ for every $A \in \mathcal{B}^2$, where $\Delta := \{(s_1, s_2) \in E^2; s_1 = s_2\}$ and $\mathrm{pr}_1 : E^2 \to E, (s_1, s_2) \mapsto s_1$ is the projection on the first coordinate. Note that this means that $\lambda_\Delta(A \times B) = \lambda(A \cap B)$ for all $A, B \in \mathcal{B}$. For the Poisson process η, we have*

(i) $\mu_1 = \lambda$;
(ii) $\mu_2 = \lambda^2 + \lambda_\Delta$
(iii) $\mu_{[2]} = \lambda^2$;
(iv) $\mathrm{cov} = \lambda_\Delta$.

Proof. (i) Follows immediately from the fact that the expectation of a Poisson random variable is its parameter.

(ii) We only have to show the equality of the measures for rectangles $A \times B \in E^2$, since the set of these rectangles is a π-system (i.e. a system of sets that is closed under

intersections) that generates $\mathcal{B}^2$. We have, by the definition of a Poisson process,

$$\begin{aligned}\mathbb{E}\big(\eta(A)\eta(B)\big) &= \mathbb{E}\big(\eta(A)\eta(B\setminus A)\big) + \mathbb{E}\big(\eta(A)\eta(A\cap B)\big)\\ &= \lambda(A)\lambda(B\setminus A) + \mathbb{E}\big(\eta(A\setminus B)\eta(A\cap B)\big) + \mathbb{E}\big(\eta(A\cap B)^2\big)\\ &= \lambda(A)\lambda(B\setminus A) + \lambda(A\setminus B)\lambda(A\cap B) + \lambda(A\cap B) + \lambda(A\cap B)^2\\ &= \lambda(A)\lambda(B) + \lambda(A\cap B)\\ &= \big(\lambda^2 + \lambda_\Delta\big)(A\times B),\end{aligned}$$

where we used that $\mathbb{E}(Z^2) = \lambda + \lambda^2$ for $Z \sim \mathrm{Po}(\lambda)$ in the third line.

(iii) Follows from (ii) by the fact that $\mu_{[2]}$ is the expectation measure of $\xi^2 - \sum_{i=1}^V \delta_{(S_i,S_i)}$, which is $\mu_2 - \lambda_\Delta$.

(iv) Follows immediately from (i) and (ii). □

Corollary 2.2.P (Integrals with respect to a Poisson process). *For any $f \in \mathfrak{F}_m$, we have*

(i) $\mathbb{E}\big(\int f(s)\,\eta(ds)\big) = \int f(s)\,\lambda(ds)$;

(ii) if λ is finite, $\mathrm{var}\big(\int f(s)\,\eta(ds)\big) = \int f^2(s)\,\lambda(ds)$.

Proof. (i) follows by Propositions 2.2.E and 2.2.O(i). (ii) follows by Corollary 2.2.F(ii) and Proposition 2.2.O(iv). □

Proposition 2.2.Q (Generating functional). *The generating functional of η is given by*

$$G_\eta(h) = \exp\Big(-\int_E \big(1-h(s)\big)\,\lambda(ds)\Big)$$

for every $h \in \mathfrak{F}_u$.

Proof. We use an extension argument. First, consider functions h of the form $h(s) := 1 - \sum_{j=1}^m (1-z_j)1_{B_j}(s)$ for $m \in \mathbb{N}$, $z_j \in [0,1]$, and $B_j \in \mathcal{B}$ bounded and pairwise disjoint. Then we obtain by Proposition 2.2.H

$$\begin{aligned}G_\eta(h) = \mathbb{E}\Big(\prod_{j=1}^m z_j^{\eta(B_j)}\Big) &= \sum_{k_1,\dots,k_m\in\mathbb{Z}_+}\Big(\prod_{j=1}^m z_j^{k_j}\Big)\mathbb{P}[\eta(B_1)=k_1,\dots,\eta(B_m)=k_m]\\ &= \sum_{k_1,\dots,k_m\in\mathbb{Z}_+}\prod_{j=1}^m\Big(e^{-\lambda(B_j)}\frac{(z_j\lambda(B_j))^{k_j}}{k_j!}\Big)\\ &= \prod_{j=1}^m\Big(e^{-\lambda(B_j)}\sum_{k\in\mathbb{Z}_+}\frac{(z_j\lambda(B_j))^{k}}{k!}\Big)\\ &= \exp\Big(-\sum_{j=1}^m(1-z_j)\lambda(B_j)\Big)\\ &= \exp\Big(-\int_E\big(1-h(s)\big)\,\lambda(ds)\Big).\end{aligned}$$

Consider then a general $h \in \mathfrak{F}_u$. Such an h can be approximated by a decreasing sequence $(h_n)_{n\in\mathbb{N}}$ of functions of the above form. By multiple applications of Levi's monotone convergence theorem, we have

$$G_\eta(h) = \lim_{n\to\infty} G_\eta(h_n) = \exp\Big(-\int_E \big(1 - h(s)\big)\ \lambda(ds)\Big).$$

□

2.2.8 Construction of new point processes from given ones

In the main part of this thesis (Chapters 5 to 7) we consider point processes that are constructed by modifying one or several given point processes by way of linear transformations, random thinnings or superpositions, and then compare the results to a Poisson process. In this subsection, we briefly introduce these point process modifications and show that the class of Poisson processes is usually closed under each of them (a partial exception is the thinning).

For each modification with given (possibly infinite-dimensional) parameter, we can easily define an associated mapping from the space of point process distributions to itself (for the superposition we assume here that the point processes superimposed are identically distributed). In all three situations the $\mathrm{Po}(\lambda Leb^D)$-distribution is obtained as a fixed point of such mappings for choices of the parameter that are typical for the corresponding modification (see Equations (2.1), (2.2), and (2.5)). Finding $\mathrm{Po}(\lambda Leb^D)$ as a fixed point may be interpeted as a manifestation of the corresponding Poisson limit theorems that we examine in Chapters 5 to 7.

Transformation of point processes

Let $\widetilde{E}$ be another lcscH equipped with its Borel σ-field $\widetilde{\mathcal{B}}$. Consider an arbitrary measurable mapping $\psi : E \to \widetilde{E}$ which respects bounded sets in the sense that preimages of bounded measurable sets are bounded. Then, for any point process ξ, the transformed process $\xi\psi^{-1}$ is again a point process, defined by $(\xi\psi^{-1})(\omega) = \xi(\omega)\psi^{-1}$ for every $\omega \in \Omega$ (where $\xi(\omega)\psi^{-1}$ denotes the image measure of $\xi(\omega)$ under ψ). If ξ has the representation

$$\sum_{i=1}^{V} \delta_{S_i},$$

then it is evident that $\xi\psi^{-1}$ has the representation

$$\sum_{i=1}^{V} \delta_{\psi(S_i)}.$$

We next show that the class of Poisson processes is closed under any measurable transformation ψ that respects bounded sets.

Proposition 2.2.R. *Let η be a Poisson process on E with intensity measure λ, and let $\psi : E \to \widetilde{E}$ be measurable and such that $\psi^{-1}(B)$ is bounded for every bounded $B \in \widetilde{\mathcal{B}}$. Then $\eta\psi^{-1}$ is a Poisson process on $\widetilde{E}$ with intensity measure $\lambda\psi^{-1}$.*

Proof. We show the properties (a) and (b) from the definition of the general Poisson process.

(a) $(\eta\psi^{-1})(B) = \eta(\psi^{-1}(B)) \sim \mathrm{Po}(\lambda(\psi^{-1}(B))) = \mathrm{Po}((\lambda\psi^{-1})(B))$ for any bounded $B \in \widetilde{\mathcal{B}}$, since $\eta \sim \mathrm{Po}(\lambda)$.

(b) If $B_1, \ldots, B_r \in \widetilde{\mathcal{B}}$ are bounded and pairwise disjoint sets, then so are their preimages $\psi^{-1}(B_1), \ldots, \psi^{-1}(B_r)$. Therefore $(\eta\psi^{-1})(B_1), \ldots, (\eta\psi^{-1})(B_r)$ are independent, again since $\eta \sim \mathrm{Po}(\lambda)$. □

In Chapter 5 we consider a special class of bijective linear transformations on $E := \widetilde{E} := \mathbb{R}^{D_1} \times \mathbb{R}^{D_2}$ with $D_1 + D_2 = D$. We define, for any $T \geq 1$ and $w(T) > 0$, the mapping $\theta_T : \mathbb{R}^{D_1} \times \mathbb{R}^{D_2} \to \mathbb{R}^{D_1} \times \mathbb{R}^{D_2}$ by $\theta_T(s,t) := \big(w(T)^{1/D_1}s, T^{-1/D_2}t\big)$. In the special case that $w(T) = T$, we write $\tilde{\theta}_T$ instead of θ_T. Note that for a homogeneous Poisson process $\eta \sim \mathrm{Po}(\lambda Leb^D)$ with $\lambda > 0$, we have by Proposition 2.2.R that

$$\eta\tilde{\theta}_T^{-1} \stackrel{\mathscr{D}}{=} \eta \quad \text{for every } T \geq 1. \tag{2.1}$$

Random thinnings of point processes

Let ξ be a point process on E, and let $p : E \to [0,1]$ be a measurable function. An *independent p-thinning* ξ_p is then given in the following way: for any realization $\sum_{i=1}^{v} \delta_{s_i}$ of ξ look at each point s_i in turn and retain it with probability $p(s_i)$ or delete it with probability $1-p(s_i)$, independently of the retention/deletion decisions for any other points. Note that multiple points at the same location are thus retained independently of one another with equal probabilities. Regard the points left over by this procedure as a realization of ξ_p.

If we allow the function p to be random as well, the class of possible thinning distributions is dramatically increased. The retention decisions then need no longer be independent given ξ. We denote such a random function by π and usually think of it as a random field $\pi = \{\pi(\cdot, s);\, s \in E\}$. Conceptually, the definition runs so that, given realizations $\sum_{i=1}^{v} \delta_{s_i}$ of ξ and p of π, independent retention decisions are made as above with the retention probabilities being $p(s_i)$ for each point s_i. However, this involves a few technical considerations (essentially, we need the random field to be measurable as a mapping $\Omega \times E \to [0,1]$), which is why we give the formal definition only in Section 6.1. We then call the resulting point process ξ_π a (general) *π-thinning* of ξ.

In the thinning context, we usually refer to ξ as "the original process" and to π as "the retention field" (or, correspondingly, to p as "the retention function"). Note that both the procedure of modifying ξ and the resulting process ξ_π are referred to as π-thinning.

The proposition below yields that the class of Poisson processes is closed under independent thinning. In Chapter 6, we combine the thinning of a point process on $E = \mathbb{R}^D$ with the contraction $\kappa_T : \mathbb{R}^D \to \mathbb{R}^D$, $s \mapsto \frac{1}{T}s$, where $T \geq 1$. Note that, for a homogeneous Poisson process $\eta \sim \mathrm{Po}(\lambda Leb^D)$ with $\lambda > 0$ and for $p_T \equiv 1/T^D$, we have by Propositions 2.2.S and 2.2.R that

$$\eta_{p_T}\kappa_T^{-1} \stackrel{\mathscr{D}}{=} \eta \quad \text{for every } T \geq 1. \tag{2.2}$$

Proposition 2.2.S. *Let η be a Poisson process on E with intensity measure λ, and let $p : E \to [0,1]$ be measurable. Then η_p is a Poisson process with intensity measure $\int_\cdot p(s)\,\lambda(ds)$.*

Proof (first version). Again the properties (a) and (b) from the definition of the general Poisson process can be shown directly.

(a) Let $B \in \mathcal{B}$ be bounded, and let $\sum_{i=1}^{V} \delta_{S_i}$ be the representation of $\eta|_B$ from Proposition 2.2.M. Define X_i as indicator random variables that are, given $V, S_1, S_2, \ldots$, independent with $\mathbb{E}X_i = p(S_i)$. We call X_i the *retention decision* for the point S_i. Note that it is important here to condition on $V, S_1, S_2, \ldots$, and not only on η, since the S_i in Proposition 2.2.M can in general not be chosen in a $\sigma(\xi)$-measurable way. Then $\sum_{i=1}^{\eta(B)} X_i \overset{\mathscr{D}}{=} \eta_p(B)$. Writing $B^{(\infty)} := \bigcup_{j\in\mathbb{Z}_+} \{j\} \times B^j$, we have for $k \in \mathbb{Z}_+$,

$$\begin{aligned}
\mathbb{P}\big[\eta_p(B) = k\big] &= \int_{B^{(\infty)}} \mathbb{P}\big[\textstyle\sum_{i=1}^{l} X_i = k \;\big|\; \eta(B) = l, S_1 = s_1, \ldots, S_l = s_l\big] \\
&\qquad \cdot \mathbb{P}\big[\eta(B) \in dl, S_1 \in ds_1, \ldots, S_l \in ds_l\big] \\
&= \sum_{l=k}^{\infty} \frac{e^{-\lambda(B)}}{l!} \int_B \ldots \int_B \sum_{\substack{e_1,\ldots,e_l\in\{0,1\} \\ \sum e_i = k}} \prod_{i=1}^{l} p(s_i)^{e_i} \big(1 - p(s_i)\big)^{1-e_i} \, \lambda(ds_1) \ldots \lambda(ds_l) \\
&= \sum_{l=k}^{\infty} \frac{e^{-\lambda(B)}}{l!} \sum_{\substack{e_1,\ldots,e_l\in\{0,1\} \\ \sum e_i = k}} \prod_{i=1}^{l} \bigg(\int_B p(s)^{e_i} \big(1 - p(s)\big)^{1-e_i} \, \lambda(ds) \bigg) \\
&= \sum_{l=k}^{\infty} \frac{e^{-\lambda(B)}}{l!} \binom{l}{k} \bigg(\int_B p(s) \, \lambda(ds) \bigg)^k \bigg(\lambda(B) - \int_B p(s) \, \lambda(ds) \bigg)^{l-k} \\
&= \frac{1}{k!} \bigg(\int_B p(s) \, \lambda(ds) \bigg)^k e^{-\lambda(B)} \sum_{l=k}^{\infty} \frac{1}{(l-k)!} \bigg(\lambda(B) - \int_B p(s) \, \lambda(ds) \bigg)^{l-k} \\
&= \frac{1}{k!} \bigg(\int_B p(s) \, \lambda(ds) \bigg)^k \exp\bigg(- \int_B p(s) \, \lambda(ds) \bigg).
\end{aligned}$$

Thus $\eta_p(B) \sim \mathrm{Po}\big(\int_B p(s)\, \lambda(ds)\big)$.

(b) Let $B_1, \ldots, B_r \in \mathcal{B}$ be bounded and pairwise disjoint sets, and let $\sum_{i=1}^{V_j} \delta_{S_{ji}}$ be representations of $\eta|_{B_j}$ for $j \in \{1, \ldots, r\}$ as in Proposition 2.2.M, which can be constructed in such a way that $\big(V_j, (S_{ji})_{i\in\mathbb{N}}\big)$, $1 \le j \le r$, are independent. Define X_{ji} as the retention decisions for the S_{ji} like in part (a), that is more precisely, given $\big(V_j, (S_{ji})_{i\in\mathbb{N}}\big)_{1\le j\le r}$, let X_{ji} be independent with $\mathbb{E}X_{ji} = p(S_{ji})$. Then it is easy to see that $\big(\sum_{i=1}^{V_1} X_{1i}, \ldots, \sum_{i=1}^{V_r} X_{ri}\big) = \big(\eta(B_1), \ldots, \eta(B_r)\big)$ is independent. The formal calculation for $r = 2$ goes

$$\begin{aligned}
&\mathbb{P}\big[\textstyle\sum_{i=1}^{V_1} X_{1i} = k_1, \sum_{i=1}^{V_2} X_{2i} = k_2\big] \\
&\quad = \mathbb{E}\Big(\mathbb{P}\Big[\textstyle\sum_{i=1}^{V_1} X_{1i} = k_1, \sum_{i=1}^{V_2} X_{2i} = k_2 \;\Big|\; V_1, (S_{1i})_{1\le i\le V_1}, V_2, (S_{2i})_{1\le i\le V_2} \Big] \Big) \\
&\quad = \mathbb{E}\Big(\mathbb{P}\big[\textstyle\sum_{i=1}^{V_1} X_{1i} = k_1 \;\big|\; V_1, (S_{1i})_{1\le i\le V_1} \big] \, \mathbb{P}\big[\textstyle\sum_{i=1}^{V_2} X_{2i} = k_2 \;\big|\; V_2, (S_{2i})_{1\le i\le V_2} \big] \Big) \\
&\quad = \mathbb{E}\Big(\mathbb{P}\big[\textstyle\sum_{i=1}^{V_1} X_{1i} = k_1 \;\big|\; V_1, (S_{1i})_{1\le i\le V_1} \big] \Big) \, \mathbb{E}\Big(\mathbb{P}\big[\textstyle\sum_{i=1}^{V_2} X_{2i} = k_2 \;\big|\; V_2, (S_{2i})_{1\le i\le V_2} \big] \Big) \\
&\quad = \mathbb{P}\big[\textstyle\sum_{i=1}^{V_1} X_{1i} = k_1 \big] \, \mathbb{P}\big[\textstyle\sum_{i=1}^{V_2} X_{2i} = k_2 \big],
\end{aligned}$$

where, for the third line, we used the conditional independence of the X_{ji} along with the fact that the conditional distribution of $(X_{ji})_{i\in\mathbb{N}}$ for some fixed $j \in \{1,\dots,r\}$ is only a function of $\big(V_j, (S_{ji})_{i\in\mathbb{N}}\big)$. For the fourth line, we used the independence of $\big(V_1, (S_{1i})_{i\in\mathbb{N}}\big)$ and $\big(V_2, (S_{2i})_{i\in\mathbb{N}}\big)$. □

The above proof is a little awkward, both for its somewhat involved formulae and its extensive use of concrete representations of η. To demonstrate the power of the generating functional approach introduced in Subsection 2.2.6, we give an alternative proof, which is more elegant.

Proof (second version, using generating functionals). By Proposition 2.2.Q, the Po(λ)-process η has the generating functional given by

$$G_\eta(h) = \exp\Big(-\int_E \big(1-h(s)\big)\,\lambda(ds)\Big) \tag{2.3}$$

for every $h \in \mathfrak{F}_u$. Furthermore, for any point process ξ on E with generating functional G_ξ its p-thinning has the generating functional given by

$$G_{\xi_p}(h) = G_\xi(ph+1-p) \tag{2.4}$$

for every $h \in \mathfrak{F}_u$. This follows, because for $\xi = \sum_{i=1}^V \delta_{S_i}$,

$$\begin{aligned} G_{\xi_p}(h) &= \mathbb{E}\Big(\prod_{i=1}^V \big(X_i h(S_i) + 1 - X_i\big)\Big) \\ &= \mathbb{E}\Big(\mathbb{E}\Big(\prod_{i=1}^V \big(X_i h(S_i) + 1 - X_i\big) \;\Big|\; \xi\Big)\Big) \\ &= \mathbb{E}\Big(\prod_{i=1}^V \big(p(S_i)h(S_i) + 1 - p(S_i)\big)\Big) \\ &= G_\xi(ph+1-p). \end{aligned}$$

Combining Equations (2.3) and (2.4) yields

$$\begin{aligned} G_{\eta_p}(h) = G_\eta(ph+1-p) &= \exp\Big(-\int_E \big(1-\big[p(s)h(s)+1-p(s)\big]\big)\,\lambda(ds)\Big) \\ &= \exp\Big(-\int_E \big(1-h(s)\big)p(s)\,\lambda(ds)\Big) \end{aligned}$$

for every $h \in \mathfrak{F}_u$, which is the generating functional of the Poisson process with intensity measure $\int_\cdot p(s)\lambda(ds)$. Thus, the statement follows by Proposition 2.2.G. □

It is easy to see that the class of Poisson processes is not closed under general thinning (i.e. with a random retention field π), not even if ξ and π are independent. Take for example the Poisson process η on $E = \{0,1\}$ with intensity measure $\lambda(\delta_0+\delta_1)$, $\lambda > 0$, so that $\eta(0) := \eta(\{0\})$ and $\eta(1) := \eta(\{1\})$ are both Po(λ)-distributed. Furthermore let $\pi(\cdot, s) =: X \sim \mathrm{Be}(p)$ for all $s \in E$, where $p \in (0,1)$ and $X \perp\!\!\!\perp \eta$. Then

$$\mathbb{E}\big(\eta_\pi(0)\eta_\pi(1)\big) = \mathbb{E}\big(\eta_\pi(0)\eta_\pi(1)\,|\,X=1\big)p = \mathbb{E}\big(\eta(0)\eta(1)\big)p = \lambda^2 p,$$

whence, by analogous calculations for $\mathbb{E}\big(\eta_\pi(0)\big)$ and $\mathbb{E}\big(\eta_\pi(1)\big)$, it follows that

$$\operatorname{cov}(\eta_\pi(0), \eta_\pi(1)) = p(1-p)\lambda^2.$$

Thus η_π is not a Poisson process, because we have two disjoint sets whose point counts are positively correlated.

In spite of this set-back, we can still generalize Proposition 2.2.S considerably if we focus on the larger class of Cox processes.

Proposition 2.2.T. *Let η be a Cox process on E with directing measure Λ, and let $\pi = \{\pi(\cdot, s);\, s \in E\}$ be a measurable random field with $\pi \perp\!\!\!\perp_\Lambda \eta$ (i.e. π is independent of η given Λ). Then η_π is a Cox process with directing measure $\int_\cdot \pi(s)\, \Lambda(ds)$.*

Proof. Because of $\pi \perp\!\!\!\perp_\Lambda \eta$, we have that

$$\mathscr{L}(\eta \,|\, \Lambda = \lambda, \pi = p) = \mathscr{L}(\eta \,|\, \Lambda = \lambda) = \mathrm{Po}(\lambda)$$

for almost every λ and almost every p. Hence

$$\mathscr{L}(\eta_\pi \,|\, \Lambda = \lambda, \pi = p) = \mathrm{Po}\bigg(\int_\cdot p(s)\ \lambda(ds)\bigg)$$

for a.e. λ and π by Proposition 2.2.S, which yields, by integration,

$$\mathscr{L}(\eta_\pi) = \mathrm{Cox}\bigg(\int_\cdot \pi(s)\ \Lambda(ds)\bigg).$$

□

Superpositions of point processes

Let $\xi_1, \dots, \xi_n$ be point processes on E. The superposition of $\xi_1, \dots, \xi_n$ is then just their sum $\sum_{k=1}^n \xi_k$ (i.e. $\big(\sum_{k=1}^n \xi_k\big)(A) = \sum_{k=1}^n \xi_k(A)$ for every $A \in \mathcal{B}$), which of course again is a point process. This definition can obviously be extended to an infinite sequence $(\xi_k)_{k\in\mathbb{N}}$ of point processes, provided that every realization of the sum $\sum_{k=1}^\infty \xi_k$ is locally finite.

Again, we obtain a result about the closedness of the class of Poisson processes. Note that the following proposition contains Proposition 2.2.K as a special case (let η_k from Proposition 2.2.U be a $\mathrm{Po}(\lambda|_{E_k})$-process on E_k and the zero measure on E_k^c).

Proposition 2.2.U. *Let $\eta_1, \eta_2, \dots$ be independent Poisson processes on E with intensity measures $\lambda_1, \lambda_2, \dots$ that satisfy $\sum_{k=1}^\infty \lambda_k \in \mathfrak{M}$. We allow $\eta_k = 0$ for any of the processes. Then $\sum_{k=1}^\infty \eta_k$ is a Poisson process with intensity measure $\sum_{k=1}^\infty \lambda_k$.*

Proof. The proof is precisely the same as the one of Proposition 2.2.K, except for the more general η_k and the obvious changes in notation. We show the properties (a) and (b) from the definition of the general Poisson process.

(a) For bounded $B \in \mathcal{B}$, $\sum_{k=1}^\infty \eta_k(B)$ is $\mathrm{Po}\big(\sum_{k=1}^\infty \lambda_k(B)\big)$-distributed by the fact that sums of finitely many independent Poisson random variables are Poisson distributed and a very simple limit argument.

(b) For bounded and pairwise disjoint $B_1, \dots, B_r \in \mathcal{B}$, the point counts $\sum_{k=1}^\infty \eta_k(B_j)$, $1 \le j \le r$, are independent, because they are functions of disjoint selections from $\eta_k(B_j)$, $k \in \mathbb{N}$, $1 \le j \le r$, which are independent. □

It is evident that the result is in general wrong if the Poisson processes superimposed are not independent, because then the superposition has usually dependent point counts.

In Chapter 7, we combine the superimposing with a mechanism that increases the sparsity of the point process distributions. Consider for example the dilation $\chi_n : \mathbb{R}^D \to \mathbb{R}^D$, $s \mapsto n^{1/D}s$, where $n \in \mathbb{N}$. Let $(\eta_k)_{k\in\mathbb{N}}$ be an i.i.d. sequence of homogeneous Poisson processes on $\mathbb{R}^D$, so that $\eta_k \sim \mathrm{Po}(\lambda Leb^D)$ for some $\lambda > 0$. We then obtain by Propositions 2.2.R and 2.2.U that

$$\sum_{k=1}^{n} \eta_k \chi_n^{-1} \stackrel{\mathscr{D}}{=} \eta \quad \text{for every } n \in \mathbb{N}. \tag{2.5}$$

2.2.9 Descriptive functions for point processes

Let $\xi \neq 0$ be a simple stationary point process on $E := \mathbb{R}^D$ equipped with the Euclidean metric. Stationarity implies that $\mathbb{E}\xi = \lambda Leb^D$ for some $\lambda > 0$. Many ways have been considered to give descriptive functions for certain characteristics of such a point process. Apart from the moment measures μ_k and $\mu_{[k]}$, which at least for $k \geq 2$ are already quite complicated objects, some of the most well-known descriptives are the distance-based functions, which are usually denoted by F, G, J, K, and L. These functions describe certain distributions associated with distances in $\mathbb{R}^D$ that involve points of the process ξ. We first define the F-function.

Definition. The *F-function* of ξ (or *empty space function*, or *spherical contact distribution function*) is the distribution function of the distance from the origin to the nearest point of ξ; that is, $F : \mathbb{R}_+ \to [0, 1]$ is defined by

$$F(r) := \mathbb{P}\big[\xi(\mathbb{B}(0, r)) \geq 1\big] \quad \text{for every } r \geq 0.$$

The term "spherical contact distribution function" will become clearer in the broader context of Subsection 2.3.6, where we define contact distribution funtions for stationary random closed sets (RACS). See Remark 2.3.G for details.

The other functions mentioned above require additional theory (Palm theory and preferably also the concept of reduced moment measures), and we do not use these functions later on. Nevertheless, as they are defined via distances in the state space, they are promising candidates for basing expressions on that can be bounded by d_2-distances. See for example the results for the F-function in Subsection 3.3.2. We briefly give the definitions and some explanations for the other functions.

G-function: also called the *nearest neighbor function*. The G-function is the distribution function of the distance from a typical point of the point process to its nearest neighbor. We have

$$G(r) := \mathbb{P}\big[\xi(\dot{\mathbb{B}}(0, r)) \geq 1 \,\big|\, \xi(\{0\}) \geq 1\big] \quad \text{for every } r \geq 0,$$

where we write $\mathbb{P}[\xi \in \cdot \,|\, \xi(\{0\}) \geq 1]$ for the Palm distribution of ξ given a point in 0 (see Kallenberg (1986), Chapter 10, for the definition).

J-function: introduced by van Lieshout and Baddeley (1996). It compares the neighborhood of a typical point of the process with the neighborhood of an arbitrary location in

space. The J-function can relatively often be computed and has a number of nice theoretical properties, such as its ability to capture the effective range of point interaction. It is given as

$$J(r) := \frac{1 - G(r)}{1 - F(r)} \quad \text{for every } r \geq 0.$$

K-function: introduced by Ripley (1976). Up to a normalizing constant, it gives the expected number of points within any given distance from a typical point of the point process. The K-function measures the amount of clustering around a typical point in a somewhat different way than the G-function. Note in particular that while the G-function sees only the nearest neighbor of the point, the K-function sees all the points within the given distance (but gives only information "in expectation"). We have

$$K(r) := \frac{1}{\lambda}\mathbb{E}\big(\xi(\dot{\mathbb{B}}(0,r)) \;\big|\; \xi(\{0\}) \geq 1\big) \quad \text{for every } r \geq 0.$$

L-function: a slightly adapted version of the K-function that is often preferred in applications. See for example the form of the L-function for a Poisson process below, which makes deviations from a Poisson assumption for small distances easier to detect. We define

$$L(r) := \sqrt[D]{\frac{K(r)}{\alpha_D}} \quad \text{for every } r \geq 0.$$

All of these functions are treated in more detail in Møller and Waagepetersen (2004), Chapter 4.

Statistical point pattern analyses usually begin with an examination of estimates of these distance-based functions, which are then often compared to the corresponding theoretical functions for a Poisson process. The latter can easily be computed as follows.

Proposition 2.2.V. *Let η be a stationary (i.e. homogeneous) Poisson process with intensity $\lambda > 0$. Then, for any $r \geq 0$,*

(i) $F(r) = 1 - e^{-\lambda\alpha_D r^D}$;
(ii) $G(r) = 1 - e^{-\lambda\alpha_D r^D}$;
(iii) $J(r) = 1$;
(iv) $K(r) = \alpha_D r^D$;
(v) $L(r) = r$.

Proof. We need the fact that the reduced Palm distribution of a Poisson process η is again the same Poisson process distribution, that is

$$\mathbb{P}\big[\eta - \delta_0 \in N \;\big|\; \eta(\{0\}) \geq 1\big] = \mathbb{P}[\eta \in N]$$

for all $N \in \mathcal{N}$. This result is usually known as Slivnyak's theorem (see e.g. Kallenberg (1986)). The expressions claimed follow now immediately from the fact that the point counts $\eta(B)$ are Poisson distributed. □

2.3 Elementary stochastic geometry

We give here a brief introduction to some basic concepts from stochastic geometry, of which we make use later on, mainly in Chapter 3 and Subsection 6.5.1. For more detailed information, the reader is referred to Stoyan, Kendall and Mecke (1987), which gives a rather applied introduction, and to Schneider and Weil (2000), which gives profound theoretical backing.

2.3.1 Random closed sets (RACS)

Denote by $\mathcal{F}$ the system of closed subsets of E and by $\mathcal{C}$ the subsystem of compact sets. Write furthermore, for any subset $A \subset E$, $\mathcal{F}^A := \{F \in \mathcal{F}; F \cap A = \emptyset\}$ for the system of closed sets missing A, and $\mathcal{F}_A := \{F \in \mathcal{F}; F \cap A \neq \emptyset\}$ for the system of closed sets hitting A. We consider the *hit-or-miss topology* on $\mathcal{F}$, which is the topology generated by

$$\{\mathcal{F}^C; C \subset E \text{ compact}\} \cup \{\mathcal{F}_G; G \subset E \text{ open}\}.$$

With this topology, $\mathcal{F}$ is an lcscH that is even compact (see e.g. Schneider and Weil (2000), Satz 2.1.2). We equip $\mathcal{F}$ with its Borel σ-algebra $\mathcal{B}(\mathcal{F})$. This σ-algebra is generated by any *one* of the following systems of sets (see Schneider and Weil (2000), Lemma 1.3.1 and the remark following it, and note that the corresponding proofs are not limited to the case $E = \mathbb{R}^D$):

$$\{\mathcal{F}^C; C \subset E \text{ compact}\}, \{\mathcal{F}_G; G \subset E \text{ open}\},$$
$$\{\mathcal{F}_C; C \subset E \text{ compact}\}, \text{ or } \{\mathcal{F}^G; G \subset E \text{ open}\}.$$

Definition. A measurable mapping $\Xi : \Omega \to \mathcal{F}$ is called a *random closed set (RACS)* in E.

Let $E := \mathbb{R}^D$. As for point processes, an important concept is stationarity. Recall that we write $s + A := \{s + a; a \in A\}$ for any $s \in \mathbb{R}^D$ and any $A \subset \mathbb{R}^D$.

Definition. A RACS Ξ in $\mathbb{R}^D$ is called *stationary* if $\mathscr{L}(s+\Xi) = \mathscr{L}(\Xi)$ for every $s \in \mathbb{R}^D$.

2.3.2 Simple point processes as RACS

Our idea of a point process is that of randomly scattered points on the space E. Since we allow only finitely many points in every bounded set, it is easily seen that every realization of a point process that has no more than one point per location in space can be interpreted as a closed subset of E. On the other hand, every realization of a RACS whose intersection with any bounded set is finite can obviously be interpreted as a locally finite point measure. It is therefore natural to ask if point processes and RACS are the same if we impose the corresponding conditions on all their realizations.

The following proposition anwers this question in the affirmative. We call a subset A of E *locally finite* if $A \cap B$ is finite for every (topologically) bounded subset B of E.

Proposition 2.3.A. *Let ξ be a point process whose realizations satisfy $\xi(\omega)(\{s\}) \leq 1$ for all $\omega \in \Omega$ and all $s \in E$. Then $\Xi_\xi := \{s \in E; \xi(\{s\}) = 1\}$ is a RACS.*
Conversely, let Ξ be a RACS whose realizations are all locally finite. Then the mapping $\xi_\Xi : \Omega \to \mathfrak{N}$ given by $\xi_\Xi(\omega)(A) := \#\big(\Xi(\omega) \cap A\big)$ is a point process on E.

Proof. Let $\omega \in \Omega$ and $s \in E \setminus \Xi_\xi(\omega)$. Since E is locally compact, there is a compact neighborhood of s, which consequently contains only finitely many points of $\xi(\omega)$. By the Hausdorff property, there is then also a neighborhood of s that contains no points of $\Xi_\xi(\omega)$, whence it follows that $\Xi_\xi(\omega) \in \mathcal{F}$. It remains to be shown that Ξ_ξ is $\mathcal{A}$-$\mathcal{B}(\mathcal{F})$-measurable. Let $C \subset E$ be compact and consider $\mathcal{F}^C$. We have

$$\Xi_\xi^{-1}(\mathcal{F}^C) = \big\{\omega; \Xi_\xi(\omega) \cap C = \emptyset\big\} = \big\{\omega; \xi(\omega)(C) = 0\big\} \in \mathcal{A},$$

because ξ is $\mathcal{A}$-$\mathcal{N}$-measurable. Since $\{\mathcal{F}^C; C \subset E \text{ compact}\}$ generates $\mathcal{B}(\mathcal{F})$, this yields the required measurability.

Conversely, it is clear that $\xi_\Xi(\omega) \in \mathfrak{N}$ for every ω. We show that ξ_Ξ is $\mathcal{A}$-$\mathcal{N}$-measurable. Choose a countable dense subset $\Sigma \subset E$. Let $G \subset E$ be bounded and open, and $k \in \mathbb{Z}_+$. We have with $\mathfrak{N}_k(G) := \{\varrho \in \mathfrak{N};\, \varrho(G) \le k\}$ that

$$\xi_\Xi^{-1}\big(\mathfrak{N}_k(G)\big) = \bigcap_{n\in\mathbb{N}} \overset{\neq}{\bigcup_{s_1,\dots,s_k\in\Sigma}} \left\{\omega;\, \Xi(\omega) \cap \left(G \setminus \left(\bigcup_{i=1}^k \mathbb{B}(s_i, 1/n)\right)\right) = \emptyset\right\} \in \mathcal{A},$$

because Ξ is $\mathcal{A}$-$\mathcal{B}(\mathcal{F})$-measurable. Note that we use the metric d from Section 2.1 for the above construction. Lemma 3.1.1 in Schneider and Weil (2000) implies that the system $\big\{\mathfrak{N}_k(G); k \in \mathbb{Z}_+, G \subset E \text{ bounded and open}\big\}$ generates $\mathcal{N}$. Thus we obtain the required measurability. □

Remark 2.3.B. Note that Proposition 2.3.A remains true if the requirements for ξ and Ξ only hold almost surely, with the following proviso: Ξ_ξ and ξ_Ξ can be defined only almost surely and therefore have to be extended in a measurable way. If the underlying probability space is complete (which can always be arranged), the extension may be chosen arbitrarily, otherwise choose the extension for example to be constant on the null set on which it has to be defined.
Thus, simple point processes and a.s. locally finite RACS are, up to $\mathbb{P}$-null sets, two models for one and the same thing.

2.3.3 Particle processes

We consider now a more interesting construction of RACS based on point processes. From Subsection 2.3.1 we know that the space $\mathcal{F}$ is an lcscH. So it is meaningful to consider point processes on $\mathcal{F}$. For our purposes it is enough to consider the smaller space $\mathcal{C}' := \mathcal{C} \setminus \{\emptyset\}$. For technical reasons (see Schneider and Weil (2000), Section 4.2), we construct our point processes in the following way.

Definition. A point process ξ on the lcscH $\mathcal{F}' := \mathcal{F} \setminus \{\emptyset\}$ that is concentrated on $\mathcal{C}'$, and whose expectation measure μ_1 exists (i.e. is locally finite) is called a *particle process*.

Note that $\mathcal{F}'$ is again an lcscH, because it is an open subset of an lcscH. The definition enables us to model RACS with the aid of the following proposition.

Proposition 2.3.C. *Let $\xi = \sum_{i=1}^V \delta_{\Xi_i}$ be a particle process. Then the union set $\Xi_\xi := \bigcup_{i=1}^V \Xi_i$ is a RACS.*

Proof. See Schneider and Weil (2000), Satz 3.5.3. The authors assume $E = \mathbb{R}^D$, but it can be easily checked that their proof remains true for general E. □

Note that, for $E = \mathbb{R}^D$, the converse direction is also true, that is every RACS in $\mathbb{R}^D$ can be written as the union set of a particle process (Schneider and Weil (2000), Satz 4.4.2).

Still for $E = \mathbb{R}^D$, we define the stationarity of a particle process in the expected way.

Definition. A particle process $\xi = \sum_{i=1}^{V} \delta_{\Xi_i}$ is called *stationary* if for any $s \in E$,

$$\mathscr{L}\Big(\sum_{i=1}^{V} \delta_{s+\Xi_i}\Big) = \mathscr{L}\Big(\sum_{i=1}^{V} \delta_{\Xi_i}\Big).$$

It is easy to see that the union set of a stationary particle process is stationary (as a RACS).

2.3.4 Boolean models

A more intuitive way of obtaining a RACS based on a point process is by attaching i.i.d. random compact sets to the points of a process on E. Note that the RACS construction in Subsection 2.3.2 can be interpreted as a very simple special case of this.

For obvious reasons, we have to set $E := \mathbb{R}^D$ in this subsection. Only the stationary case is considered, because it is all that we need in later chapters, and the technical condition that ensures compatibility with particle processes is somewhat easier in this case. Write $A + B := \{a + b;\, a \in A, b \in B\}$ for the (Minkowski-)sum of sets $A, B \subset \mathbb{R}^D$.

Definition. Let $\xi = \sum_{i=1}^{V} \delta_{S_i}$ be a stationary point process on $\mathbb{R}^D$, and $\Xi_1, \ldots, \Xi_n$ random elements in $\mathcal{C}'$ that are i.i.d., independent also of ξ, and satisfy

$$\mathbb{E}\Big(Leb^D\big(\Xi_i + \mathbb{B}(0,r)\big)\Big) < \infty \qquad \text{for some } r > 0. \tag{2.6}$$

Then

$$\Xi := \bigcup_{i=1}^{V} \big(S_i + \Xi_i\big)$$

is called a stationary *germ-grain model*. The process ξ is called the *germ-process*, and the random compact sets Ξ_i are called *grains*. We call an arbitrary random set Ξ_0 that has the same distribution as the Ξ_i a *typical grain*.

Strictly speaking, a germ-grain model as defined above is not a RACS, as it may have realizations that are not closed subsets of $\mathbb{R}^D$. However, by Condition (2.6) it can be seen that after modification on a $\mathbb{P}$-null set $\zeta := \sum_{i=1}^{V} \delta_{S_i+\Xi_i}$ is a particle process, and hence, by Proposition 2.3.C, its union set $\bigcup_{i=1}^{V} \big(S_i + \Xi_i\big)$ is a RACS. In what follows we always tacitly assume that Ξ is modified in such a way that it is a RACS.

A very important germ-grain model is the Boolean model. For this RACS a number of characteristic properties can be explicitly calculated (see Subsections 2.3.5 and 2.3.6).

Definition. Let $\lambda \in \mathbb{R}_+$. A stationary germ-grain model Ξ is called stationary *Boolean model* (or *Poisson germ-grain model*) with intensity λ if its germ process ξ is a $\mathrm{Po}(\lambda Leb^D)$-process.

The next proposition shows that stationary Boolean models and union sets of stationary Poisson particle processes are one and the same thing. Note that for the non-stationary analogs this would no longer be true: not every union set of a Poisson particle process can be obtained by attaching i.i.d. sets *independently* to the points of a Poisson process. In order to remove the redundancy in the representation of the germ-grain model, we write $c(C)$ for the center of the (unique!) circumsphere of the compact set $C \in \mathcal{C}'$, and $\mathcal{C}_0 := \{C \in \mathcal{C}'; c(C) = 0\}$. We then restrict the distribution of the typical grain to $\mathcal{C}_0$, so that we have a strict separation between the position of the i-th grain (given by the germ S_i it is attached to) and its shape.

Proposition 2.3.D. *The class of distributions of stationary Boolean models is equal to the class of distributions of union sets of stationary Poisson particle processes. The relation between the intensity λ of the germ process and the distribution of the typical grain Ξ_0 on $\mathcal{C}_0$ on the one hand, and the intensity measure μ_1 of the Poisson particle process on the other hand, is given by*

$$\mu_1(\mathcal{G}) = \lambda \mathbb{E}\Big(\int_{\mathbb{R}^D} 1_{\mathcal{G}}(s + \Xi_0)\, Leb^D(ds) \Big) \tag{2.7}$$

for any measurable $\mathcal{G} \subset \mathcal{C}'$. Note that $(\lambda, \mathscr{L}(\Xi_0))$ and μ_1 determine each other completely for $\mu_1 \neq 0$.

Proof. See Schneider and Weil (2000), Satz 4.4.3. The relation between λ, $\mathscr{L}(\Xi_0)$, and μ_1 follows by Satz 4.2.2. □

2.3.5 The capacity functional

Definition. Let Ξ be a RACS in E. The functional $T_\Xi : \mathcal{C} \to [0,1]$ that is given by

$$T_\Xi(C) := \mathbb{P}[\Xi \cap C \neq \emptyset] \quad \text{for every } C \in \mathcal{C}$$

is called the *capacity functional* of Ξ.

Capacity functionals for RACS play the role of distribution functions for real-valued random variables. In particular, the distribution of a RACS is characterized by its capacity functional. The celebrated Choquet theorem characterizes the functionals $T : \mathcal{C} \to [0,1]$ for which there is a RACS Ξ with capacity functional T. See Schneider and Weil (2000), Section 2.2 for details.

The capacity functional of a stationary Boolean model can be easily computed by the characterization given in Proposition 2.3.D.

Lemma 2.3.E. *Let Ξ be a stationary Boolean model with intensity $\lambda > 0$ and typical grain Ξ_0. Then its capacity functional is given by*

$$T_\Xi(C) = 1 - \exp\Big(-\lambda \mathbb{E}\big(Leb^D(\Xi_0^* + C) \big) \Big) \quad \textit{for every } C \in \mathcal{C},$$

where for any set $A \subset \mathbb{R}^D$ we denote by $A^* := \{-s, s \in A\}$ its reflection in the origin.

Proof. Let μ_1 be the expectation measure of the Poisson particle process ζ whose union set is Ξ, that is μ_1 is given by Equation (2.7). We then have

$$T_\Xi(C) = \mathbb{P}[\zeta(\mathcal{F}_C) \geq 1] = 1 - \exp\bigl(-\mu_1(\mathcal{F}_C)\bigr).$$

Note that $(s + A) \cap C \neq \emptyset$ is equivalent to $s \in (A^* + C)$ for any sets $A, C \subset \mathbb{R}^D$. Thus, by Equation (2.7),

$$\begin{aligned}\mu_1(\mathcal{F}_C) &= \lambda\mathbb{E}\biggl(\int_{\mathbb{R}^D} 1_{\mathcal{F}_C}(s + \Xi_0)\; Leb^D(ds)\biggr)\\ &= \lambda\mathbb{E}\biggl(\int_{\mathbb{R}^D} 1_{\Xi_0^* + C}(s)\; Leb^D(ds)\biggr)\\ &= \lambda\mathbb{E}\bigl(Leb^D(\Xi_0^* + C)\bigr)\end{aligned}$$

for every $C \in \mathcal{C}$, which yields the statement. □

2.3.6 Descriptive functions for stationary RACS

We introduce here the volume fraction and the contact distribution function of a stationary RACS Ξ. The former gives the expected volume of Ξ per unit area, and hence is descriptive of the (relative) size of the RACS. The latter gives, for a fixed compact set C, the distribution function of the minimal scaling factor necessary for C to hit Ξ given the center of the scaling lies within C but not in Ξ. Hence it is able to capture also the shape aspect of the RACS, in addition to its size.

The volume fraction

Definition. Let Ξ be a stationary RACS in $\mathbb{R}^D$. We call

$$p_\Xi := \mathbb{E}\bigl(Leb^D(\Xi \cap [0,1]^D)\bigr)$$

the *volume fraction* (or *volume density*) of Ξ.

The volume fraction is just the value of the capacity functional at $\{0\}$.

Proposition 2.3.F. *Let Ξ be a stationary RACS in $\mathbb{R}^D$ with capacity functional T_Ξ. Then its volume fraction satisfies*

$$p_\Xi = \mathbb{P}[0 \in \Xi] = T_\Xi(\{0\}),$$

and furthermore

$$p_\Xi = \frac{\mathbb{E}\bigl(Leb^D(\Xi \cap B)\bigr)}{Leb^D(B)}$$

for every Borel set $B \subset \mathbb{R}^D$ with $0 < Leb^D(B) < \infty$.

Proof. See Schneider and Weil (2000), Satz 1.4.6, and the subsequent remark. □

By the first statement we obtain immediately from Lemma 2.3.E that, if Ξ is a Boolean model with intensity λ and typical grain Ξ_0, then

$$p_\Xi = 1 - \exp\Bigl(-\lambda\mathbb{E}\bigl(Leb^D(\Xi_0)\bigr)\Bigr). \tag{2.8}$$

The contact distribution function

Definition. Let Ξ be a stationary RACS in $\mathbb{R}^D$ with volume fraction $p_\Xi < 1$, and let C be a compact set containing the origin. We call the function $H^{(C)} := \mathbb{R}_+ \to [0,1]$ that is given by

$$H^{(C)}(r) := \mathbb{P}\big[rC \cap \Xi \neq \emptyset \,|\, 0 \notin \Xi\big]$$

the *contact distribution function* of Ξ with structuring element C, where we set $rC := \{rs;\, s \in C\}$. If C is the Euclidean unit ball $\mathbb{B}(0,1)$, then $H^{(C)} =: H_s$ is called the *spherical contact distribution function*; if C is the unit line segment $[\![0,u]\!]$, i.e. the convex hull of $\{0,u\}$, where $u \in \mathbb{R}^D$ with $|u| = 1$, then $H^{(C)} =: H_l^{(u)}$ is called a *linear contact distribution function.*

Again we can express the function via the capacity functional, for example as

$$H^{(C)}(r) = 1 - \frac{1 - T_\Xi(rC)}{1 - T_\Xi(\{0\})},$$

whence we obtain for the Boolean model with intensity λ and typical grain Ξ_0 that

$$H^{(C)}(r) = 1 - \exp\Big(-\lambda\big[\mathbb{E}\big(Leb^D(\Xi_0^* + rC)\big) - \mathbb{E}\big(Leb^D(\Xi_0)\big)\big]\Big).$$

Remark 2.3.G. Note that, for the RACS Ξ_ξ given (almost surely) as the set of points of the simple stationary point process ξ (see Remark 2.3.B), the spherical contact distribution function H_s is equal to the F-function of ξ, because $\mathbb{P}\big[\mathbb{B}(0,r) \cap \Xi_\xi \neq \emptyset\big] = \mathbb{P}\big[\xi(\mathbb{B}(0,r)) \geq 1\big]$ and $\mathbb{P}[0 \notin \Xi_\xi] = 1$.

2.4 Probability metrics

In this section, we give an overview of the properties and interrelations of various well-known probability metrics that are used in this thesis. By a probability metric we mean any proper metric (i.e. with values in $\mathbb{R}_+$) on a set of probability measures on some space Γ. In the terminology of Rachev (1991), these are the "simple" ($\mathbb{R}_+$-valued) probability metrics. We use the term "distance" for the evaluation of a metric at two specific elements.

It is the goal of this thesis to give upper bounds for distances between point process distributions in well-known approximation settings. The metric that we use for this purpose is the Barbour-Brown metric d_2, which we define essentially as two Wasserstein metrics built on top of one another. This and other metrics that are more specific to the point process setting are defined in Chapter 3, whereas in what follows, we consider more general concepts of probability metrics.

2.4.1 The total variation metric

Let $(\Gamma, \mathcal{G})$ be an arbitrary measurable space. The *total variation distance* between two probability measures ν_1 and ν_2 on Γ is defined as

$$d_{TV}(\nu_1, \nu_2) := \sup_{A \in \mathcal{G}} \big|\nu_1(A) - \nu_2(A)\big|,$$

which by an approximation argument can be written in the equivalent form

$$d_{TV}(\nu_1, \nu_2) = \sup_{f \in \mathfrak{F}_{TV}} \left| \int f \, d\nu_1 - \int f \, d\nu_2 \right|, \tag{2.9}$$

where $\mathfrak{F}_{TV} := \{f : \Gamma \to [0,1] \text{ measurable}\}$.

It is easily seen that d_{TV} is a metric (called the *total variation metric*) that is bounded by 1. The following dual formulation is very useful.

Theorem 2.4.A. *Suppose that Γ is a separable metric space and $\mathcal{G}$ its Borel σ-algebra. Then, for probability measures ν_1 and ν_2 on Γ,*

$$d_{TV}(\nu_1, \nu_2) = \min_{\substack{X_1 \sim \nu_1 \\ X_2 \sim \nu_2}} \mathbb{P}[X_1 \neq X_2].$$

Proof. See Barbour, Holst, and Janson (1992), Appendix A.1. The conditions for $(\Gamma, \mathcal{G})$ are used to ensure that the diagonal $\Delta := \{(x_1, x_2) \in \Gamma^2; x_1 = x_2\}$ is in $\mathcal{G}^2$. □

2.4.2 The Wasserstein and bounded Wasserstein metrics

Suppose that (Γ, d) is a separable metric space equipped with its Borel σ-algebra $\mathcal{G}$. Let

$$\mathfrak{F}_W := \left\{ f : \Gamma \to \mathbb{R} \,;\; |f(x) - f(y)| \leq d(x,y) \text{ for all } x, y \in \Gamma \right\} \text{ and}$$
$$\mathfrak{F}_{BW} := \left\{ f : \Gamma \to [0,1] \,;\; |f(x) - f(y)| \leq d(x,y) \text{ for all } x, y \in \Gamma \right\}$$

be the sets of 1-Lipschitz continuous functions with values in $\mathbb{R}$ and $[0,1]$, respectively. The *Wasserstein distance* and the *bounded Wasserstein distance* between two probability measures ν_1 and ν_2 on Γ are then defined as

$$d_W(\nu_1, \nu_2) := \sup_{f \in \mathfrak{F}_W} \left| \int f \, d\nu_1 - \int f \, d\nu_2 \right|$$

and

$$d_{BW}(\nu_1, \nu_2) := \sup_{f \in \mathfrak{F}_{BW}} \left| \int f \, d\nu_1 - \int f \, d\nu_2 \right|,$$

respectively, where we tacitly assume (here and later) that, for the Wasserstein distance, the measures ν_1 and ν_2 are such that $\int d(x, x_0) \, \nu_i(dx) < \infty$ for some (and hence every) $x_0 \in \Gamma$ and for $i = 1, 2$.

It can be seen that d_W is a metric on the set of such measures, and that d_{BW} is a metric on the set of all probability measures on Γ that is bounded by 1 (see the proof of Proposition 11.3.2 in Dudley (1989) for the one part of the definition of a metric that is not obvious here). Furthermore, it is easy to see that the bounded Wasserstein metric with respect to the metric d on Γ is equal to the Wasserstein metric with respect to $d \wedge 1$, which allows one to obtain results for the bounded Wasserstein metric from results for the Wasserstein metric. For the Wasserstein metric, we again have a useful dual formulation.

Theorem 2.4.B (Kantorovich-Rubinstein). *For probability measures ν_1 and ν_2 on Γ (always with $\int d(x, x_0)\, \nu_i(dx) < \infty$ for some x_0), we have*

$$d_W(\nu_1, \nu_2) = \inf_{\substack{X_1 \sim \nu_1 \\ X_2 \sim \nu_2}} \mathbb{E} d(X_1, X_2).$$

The infimum is attained if ν_1 and ν_2 are tight (i.e. for every $\varepsilon > 0$ there is $C \subset \Gamma$ compact such that $\nu_i(C) > 1 - \varepsilon$), thus for example if Γ is complete.

Proof. See Dudley (1989), Theorem 11.8.2. □

2.4.3 The Prohorov metric

Suppose that (Γ, d) is again a separable metric space equipped with its Borel σ-algebra $\mathcal{G}$. The *Prohorov distance* between two probability measures ν_1 and ν_2 on Γ is defined as

$$d_P(\nu_1, \nu_2) = \inf\big\{\varepsilon > 0; \nu_1(A) \le \nu_2(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{G}\big\},$$

where $A^\varepsilon := \{y \in \Gamma; d(y, x) < \varepsilon \text{ for some } x \in A\}$ is the open ε-halo set of A.

It can be seen that d_P is a metric that is bounded by 1 (see Dudley (1989), Theorem 11.3.1). Again there is a dual formulation. In order to state it, we first define the Ky Fan distance between Γ-valued random elements X_1 and X_2 as

$$\kappa(X_1, X_2) := \inf\big\{\varepsilon \ge 0; \mathbb{P}[d(X_1, X_2) > \varepsilon] \le \varepsilon\big\}.$$

It can be seen that κ is a metric on the space $L_0(\Omega, \Gamma)$ of equivalence classes of such random elements (w.r.t. to almost sure equality) which is furthermore bounded by 1. See Dudley (1989), Theorem 9.2.2.

Theorem 2.4.C (Strassen). *For probability measures ν_1 and ν_2 on Γ we have*

$$d_P(\nu_1, \nu_2) = \inf_{\substack{X_1 \sim \nu_1 \\ X_2 \sim \nu_2}} \kappa(X_1, X_2).$$

The infimum is attained if ν_1 and ν_2 are tight, thus for example if Γ is complete.

Proof. See Dudley (1989), Corollary 11.6.4. □

2.4.4 Common properties and interrelations

In what follows we assume always that (Γ, d) is a separable metric space equipped with its Borel σ-algebra $\mathcal{G}$. We are mainly interested in two questions. First, how are the above metrics comparable with each other (if at all), and secondly, how does convergence in any of the various metrics relate to the weak convergence of probability measures on Γ.

The answer to the first question is summarized in Figure 2.4.1. The notation and many of the results are taken from Gibbs and Su (2002), which provides a very nice overview of some of the most important probability metrics. All of the inequalities can be proved quite directly with the above definitions and results. The proofs are given as follows; note that we always require that $\overline{d} := \sup_{x,y \in \Gamma} d(x, y) \ge 1$ and $\underline{d} := \inf_{x,y \in \Gamma, x \ne y} d(x, y) \le 1$.

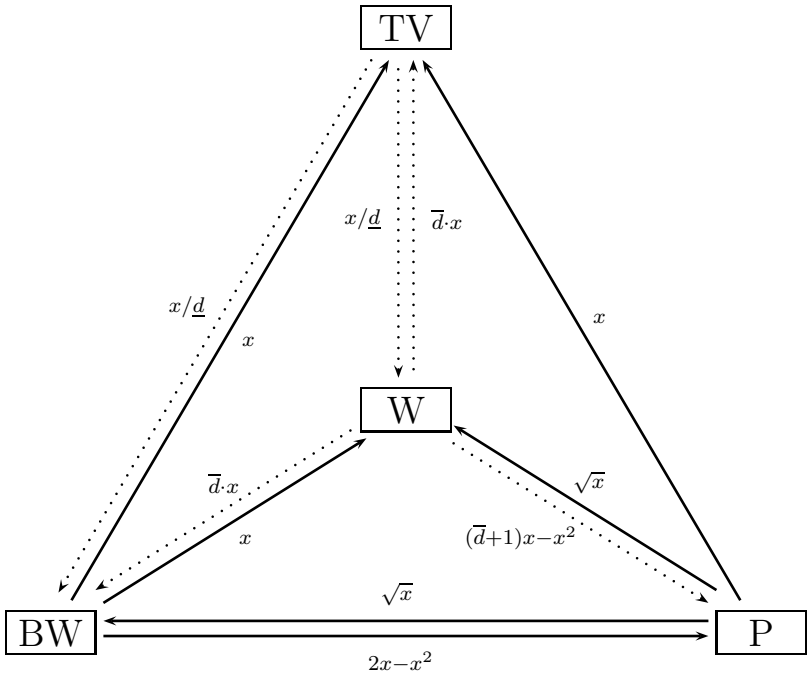

FIGURE 2.4.1: RELATIONSHIPS AMONG THE DEFINED METRICS. An arrow from A to B annotated by $f(x)$ means that $d_A \le f(d_B)$. Let $\overline{d} := \operatorname{diam}(\Gamma) := \sup_{x,y\in\Gamma} d(x,y) \ge 1$ denote the diameter of Γ. Bounds involving it are only useful if it is finite. Furthermore let $\underline{d} := \inf_{x,y\in\Gamma, x\neq y} d(x,y) \le 1$ be the minimal distance in Γ. Bounds involving it are only useful if it is positive.

$\boldsymbol{d_{BW} \le d_{TV}}$: Follows immediately from Equation (2.9).

$\boldsymbol{d_{TV} \le d_{BW}/\underline{d}}$: Assume $\underline{d} > 0$, and let $f \in \mathfrak{F}_{TV}$. Then $\underline{d}f \in \mathfrak{F}_{BW}$, since

$$|\underline{d}f(x) - \underline{d}f(y)| \le \underline{d} \le d(x,y)$$

for all $x, y \in \Gamma$ with $x \neq y$. Therefore,

$$d_{TV}(\nu_1,\nu_2) = \sup_{f\in\mathfrak{F}_{TV}} \left| \int f d\nu_1 - \int f d\nu_2 \right| \le \frac{1}{\underline{d}} \sup_{f\in\mathfrak{F}_{BW}} \left| \int f d\nu_1 - \int f d\nu_2 \right| = \frac{1}{\underline{d}} d_{BW}(\nu_1,\nu_2).$$

$\boldsymbol{d_W \le \overline{d} \cdot d_{TV}}$: Assume $\overline{d} < \infty$, and let $f \in \mathfrak{F}_W$. Then f is measurable and satisfies $\sup f(\Gamma) - \inf f(\Gamma) \le \overline{d}$, the latter because

$$|f(x) - f(y)| \le d(x,y) \le \overline{d}$$

for all $x, y \in \Gamma$. In view of the supremum in the definition of d_W, we then may assume without loss of generality that $f(\Gamma) \subset [0, \overline{d}]$, since $\nu_1\big(f - \inf f(\Gamma)\big) - \nu_2\big(f - \inf f(\Gamma)\big) = \nu_1(f) - \nu_2(f)$. Hence $f/\overline{d} \in \mathfrak{F}_{TV}$, and thus

$$d_W(\nu_1,\nu_2) = \sup_{f\in\mathfrak{F}_W} \left| \int f d\nu_1 - \int f d\nu_2 \right| \le \overline{d} \sup_{f\in\mathfrak{F}_{TV}} \left| \int f d\nu_1 - \int f d\nu_2 \right| = \overline{d} \cdot d_{TV}(\nu_1,\nu_2).$$

$\boldsymbol{d_{TV} \le d_W/\underline{d}}$: Follows from $d_{TV} \le d_{BW}/\underline{d}$ and $d_{BW} \le d_W$ below.

$\boldsymbol{d_P \le d_{TV}}$: We have for $A \in \mathcal{G}$, and for every $\varepsilon > 0$ with $|\nu_1(A) - \nu_2(A)| \le \varepsilon$, that $\nu_1(A) \le \nu_2(A) + \varepsilon$, and hence also $\nu_1(A) \le \nu_2(A^\varepsilon) + \varepsilon$. Therefore,

$$\begin{aligned} d_P(\nu_1, \nu_2) &= \inf\bigl\{\varepsilon > 0;\, \nu_1(A) \le \nu_2(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{G}\bigr\} \\ &\le \inf\bigl\{\varepsilon > 0;\, |\nu_1(A) - \nu_2(A)| \le \varepsilon \text{ for all } A \in \mathcal{G}\bigr\} = d_{TV}(\nu_1, \nu_2). \end{aligned}$$

$\boldsymbol{d_{BW} \le d_W}$: Follows immediately from the definitions.

$\boldsymbol{d_W \le \overline{d} \cdot d_{BW}}$: Assume $\overline{d} < \infty$ and let $f \in \mathfrak{F}_W$. As seen in the part "$d_W \le \overline{d} \cdot d_{TV}$", we can assume without loss of generality that $f(\Gamma) \subset [0, \overline{d}]$. Then $f/\overline{d} \in \mathfrak{F}_{BW}$, and therefore $d_W(\nu_1, \nu_2) \le \overline{d} \cdot d_{BW}(\nu_1, \nu_2)$.

$\boldsymbol{d_P^2 \le d_W}$: Follows from $d_P^2 \le d_{BW}$ below and $d_{BW} \le d_W$.

$\boldsymbol{d_W \le (\overline{d} + 1)d_P - d_P^2}$: We use a slight generalization of the proof of Theorem 3.1 in Brown and Xia (1995a). Assume $\overline{d} < \infty$. Let $X_1 \sim \nu_1$ and $X_2 \sim \nu_2$ arbitrary. For every $\varepsilon \ge 0$ with $\mathbb{P}[d(X_1, X_2) > \varepsilon] \le \varepsilon$, we have

$$\begin{aligned} \mathbb{E}d(X_1, X_2) &= \mathbb{E}\bigl(1_{\{d(X_1,X_2)>\varepsilon\}} d(X_1, X_2)\bigr) + \mathbb{E}\bigl(1_{\{d(X_1,X_2)\le\varepsilon\}} d(X_1, X_2)\bigr) \\ &\le \overline{d}\,\mathbb{P}[d(X_1, X_2) > \varepsilon] + \varepsilon\bigl(1 - \mathbb{P}[d(X_1, X_2) > \varepsilon]\bigr) \\ &\le (\overline{d} + 1)\varepsilon - \varepsilon^2. \end{aligned}$$

Taking the infimum over all $\varepsilon \ge 0$ with $\mathbb{P}[d(X_1, X_2) > \varepsilon] \le \varepsilon$ and over all $X_1 \sim \nu_1$ and $X_2 \sim \nu_2$, yields

$$d_W(\nu_1, \nu_2) \le (\overline{d} + 1)d_P(\nu_1, \nu_2) - d_P^2(\nu_1, \nu_2). \tag{2.10}$$

$\boldsymbol{d_P^2 \le d_{BW}}$: Let $A \in \mathcal{G}$, and set $f_\varepsilon(x) := (1 - d(x, A)/\varepsilon) \vee 0$ for every $x \in \Gamma$ and every $\varepsilon \in (0, 1]$, so that $\varepsilon f_\varepsilon \in \mathfrak{F}_{BW}$. Thus, we have

$$\nu_1(A) \le \int f_\varepsilon \, d\nu_1 \le \int f_\varepsilon \, d\nu_2 + \left| \int f_\varepsilon \, d\nu_1 - \int f_\varepsilon \, d\nu_2 \right| \le \nu_2(A^\varepsilon) + \frac{1}{\varepsilon} d_{BW}(\nu_1, \nu_2)$$

for every $\varepsilon \in (0, 1]$. For $d_{BW}(\nu_1, \nu_2) > 0$ set $\varepsilon = \sqrt{d_{BW}(\nu_1, \nu_2)}$ to obtain the required result; for $d_{BW}(\nu_1, \nu_2) = 0$ the result follows directly from $\nu_1(A) \le \nu_2(A^\varepsilon)$ for every $\varepsilon \in (0, 1]$ and every $A \in \mathcal{G}$.

$\boldsymbol{d_{BW} \le 2d_P - d_P^2}$: As mentioned in Subsection 2.4.2, d_{BW} with respect to d is the Wasserstein distance on Γ with respect to $d \wedge 1$. Equation (2.10) therefore yields immediately that

$$d_{BW}(\nu_1, \nu_2) \le 2d_P(\nu_1, \nu_2) - d_P^2(\nu_1, \nu_2).$$

We now turn to the second question, about the relation of the four metrics to the weak convergence of probability measures; this convergence is denoted by $\nu_n \xrightarrow{w} \nu$

for probability measures $\nu, \nu_1, \nu_2, \ldots$ on Γ. The answer is summarized in the following theorem.

Theorem 2.4.D. *Let $\nu, \nu_1, \nu_2, \ldots$ be probability measures on Γ.*
I. The following statements are equivalent.

(i) $\nu_n \xrightarrow{w} \nu$;

(ii) $d_{BW}(\nu_n, \nu) \longrightarrow 0$;

(iii) $d_P(\nu_n, \nu) \longrightarrow 0$.

II. If either $d_W(\nu_n, \nu) \longrightarrow 0$ or $d_{TV}(\nu_n, \nu) \longrightarrow 0$, we have that $\nu_n \xrightarrow{w} \nu$, but the converse statement is in general not true.

Proof. I. The equivalence of (ii) and (iii) follows immediately from the above inequalities between d_{BW} and d_P. For the equivalence to the weak convergence, see Dudley (1989), Theorem 11.3.3.

II. The convergence $\nu_n \xrightarrow{w} \nu$ follows directly from Part I and $d_{BW} \leq d_W$ resp. $d_{BW} \leq d_{TV}$. To see that the converse implications are in general not true, set $\Gamma := \mathbb{R}$, take for d the Euclidean metric, and let $\nu_n := \frac{1}{n}\delta_n + \frac{(n-1)}{n}\delta_0$, $\tilde{\nu}_n := \delta_{1/n}$, and $\nu = \delta_0$. Then $\nu_n \xrightarrow{w} \nu$ and $\tilde{\nu}_n \xrightarrow{w} \nu$, but $d_W(\nu_n, \nu) = d_{TV}(\tilde{\nu}_n, \nu) = 1$ for all n. □

Chapter 3

The Barbour-Brown metric on the space of point process distributions

When dealing with distances between point process distributions, we always assume that the corresponding point processes are defined on the compact space (E', d_0) introduced in Section 2.1. We use $(\mathfrak{M}', \mathcal{M}')$ and $(\mathfrak{N}', \mathcal{N}')$ to denote the spaces of finite measures and finite point measures on E', respectively.

In order to measure distances between point process distributions, we have to choose a metric on the space of probability distributions on $\mathfrak{N}'$, which we denote by $\mathfrak{P}(\mathfrak{N}')$. This metric is in later chapters always the Wasserstein metric d_2, defined in Section 3.2. We refer to it as the Barbour-Brown metric, according to its introduction in Barbour and Brown (1992). The d_2-metric performs in various aspects better than other possible metrics. For example, it is more appropriate than the total variation metric d_{TV} in view of the usual topology of weak convergence on $\mathfrak{P}(\mathfrak{N}')$, and it is easier to handle than the Prohorov metric d_P.

In order to define a Wasserstein metric on $\mathfrak{P}(\mathfrak{N}')$, we need to begin by determining an appropriate metric d_1 on the underlying space $\mathfrak{N}'$. This is the first topic to be dealt with in this chapter (Section 3.1). In Section 3.2, the Barbour-Brown metric is formally defined and some important properties are given. Finally, in Section 3.3, we demonstrate a number of consequences obtainable from upper bounds on the d_2-distance between two point process distributions; these are important for applying the estimation results of Chapters 5 to 7.

3.1 Metrics on the space $\mathfrak{N}'$ of finite point measures on E'

3.1.1 Definitions, properties, and interrelations

Based on the considerations in Section 2.4, we define several metrics on $\mathfrak{N}'$. In the course of this section, we drop these metrics one by one until only the relative Wasserstein metric is left, which is the one best suited to our purposes (followed by the Prohorov metric, which we also sometimes make use of). We usually denote the relative Wasserstein metric by d_1, in order to express its special position for us as the metric on $\mathfrak{N}'$ that provides the

missing link between d_0 on E' and d_2 on $\mathfrak{P}(\mathfrak{N}')$.

The definitions of the metrics are conceived as follows. For all the metrics, we always set distances between finite point measures on E' that do not have the same numbers of points equal to 1. This is done for the sake of simplicity and reflects our wish to rebuild the vague topology $\mathcal{T}_{\mathfrak{N}'}$ on $\mathfrak{N}'$, because for a sequence $(\varrho_n)_n$ of finite point measures *on a compact space* to converge vaguely to some point measure ϱ, it is necessary that ϱ_n ultimately have the same number of points as ϱ (see Theorem 3.1.B).

If the point measures $\varrho_1, \varrho_2 \in \mathfrak{N}'$ do have the same numbers of points, we define the distance between them as the corresponding total variation, Wasserstein or Prohorov distance from Section 2.4 between their normalizations $\varrho_1/|\varrho_1|$ and $\varrho_2/|\varrho_2|$ (note that the bounded Wasserstein distance would be the same as the Wasserstein distance, since $d_0 \leq 1$). Another approach is to simply extend the definitions from Section 2.4 to general finite measures with fixed total mass. For the total variation and Wasserstein metrics, it is not interesting to do so, because this results only in a multiplication of the distance by the total number of points in each of the point measures, which is insignificant with regard to the topologies generated, and is otherwise rather annoying. For the Prohorov metric, this extended form is more interesting, because it yields an intuitive interpretation of the metric (see Lemma 3.1.A) and it is widely used in the literature (note, however, that in the definition below, we trim the Prohorov metric as usually defined in the literature at 1).

We now give the formal definitions.

Definition. Let $\varrho_1, \varrho_2 \in \mathfrak{N}'$. Define the *relative total variation metric* $d_{\widetilde{TV}}$, the *relative Wasserstein metric* $d_1 := d_{\widetilde{W}}$, the *relative Prohorov metric* $d_{\widetilde{P}}$, and the *Prohorov metric* d_P by setting the corresponding distances, for $|\varrho_1| = |\varrho_2| \geq 1$, to be

(a) $d_{\widetilde{TV}}(\varrho_1, \varrho_2) := d_{TV}(\varrho_1/|\varrho_1|, \varrho_2/|\varrho_2|)$;
(b) $d_1(\varrho_1, \varrho_2) := d_{\widetilde{W}}(\varrho_1, \varrho_2) := d_W(\varrho_1/|\varrho_1|, \varrho_2/|\varrho_2|) = d_{BW}(\varrho_1/|\varrho_1|, \varrho_2/|\varrho_2|)$;
(c) $d_{\widetilde{P}}(\varrho_1, \varrho_2) := d_P(\varrho_1/|\varrho_1|, \varrho_2/|\varrho_2|)$;
(d) $d_P(\varrho_1, \varrho_2) := \inf\{\varepsilon > 0; \varrho_1(A) \leq \varrho_2(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{B}'\}$.

Set the distances equal to 1 if $|\varrho_1| \neq |\varrho_2|$, and equal to zero if $|\varrho_1| = |\varrho_2| = 0$.

It is easily seen that the mappings defined are metrics which are bounded by 1. This follows from the fact that d_{TV}, d_W, and d_P for probability measures are metrics of this kind (see Section 2.4). The extended version of the Prohorov metric defined above is bounded by 1 because $A^\varepsilon = E'$ for $\varepsilon > 1$.

Since our measures are integer valued, we can obtain much more instructive forms for three of the metrics.

Lemma 3.1.A. *For $\varrho_1, \varrho_2 \in \mathfrak{N}'$ with $|\varrho_1| = |\varrho_2| = v \geq 1$ and representations $\sum_{i=1}^v \delta_{s_{1,i}}$ and $\sum_{i=1}^v \delta_{s_{2,i}}$, respectively, we have*

(i) $d_{\widetilde{TV}}(\varrho_1, \varrho_2) = \min_{\pi \in \Sigma_v} \frac{1}{v} \sum_{i=1}^v \mathrm{I}[s_{1,i} \neq s_{2,\pi(i)}]$;
(ii) $d_1(\varrho_1, \varrho_2) = \min_{\pi \in \Sigma_v} \frac{1}{v} \sum_{i=1}^v d_0(s_{1,i}, s_{2,\pi(i)})$;
(iii) $d_P(\varrho_1, \varrho_2) = \min_{\pi \in \Sigma_v} \max_{1 \leq i \leq v} d_0(s_{1,i}, s_{2,\pi(i)})$;

where Σ_v denotes the set of all permutations of $\{1, 2, \ldots, v\}$.

Proof. (i) This follows directly from (ii), because $d_{\widetilde{TV}} = d_1$ if we choose as metric d_0 on E' the discrete metric given by $d_0(s_1, s_2) := \mathrm{I}[s_1 \neq s_2]$ for all $s_1, s_2 \in E'$.

(ii) By the Kantorovich-Rubinstein theorem (Theorem 2.4.B) we have

$$d_1(\varrho_1, \varrho_2) = d_W(\varrho_1/|\varrho_1|, \varrho_2/|\varrho_2|) = \min_{\substack{X_1 \sim \varrho_1/|\varrho_1| \\ X_2 \sim \varrho_2/|\varrho_2|}} \mathbb{E}\, d_0(X_1, X_2). \tag{3.1}$$

Setting $d_{ij} := d_0(s_{1,i}, s_{2,j})$ for all $i, j \in \{1, \dots, v\}$, this minimization is clearly equivalent to the problem

$$\begin{aligned} &\text{minimize } \frac{1}{v} \sum_{i,j=1}^{v} d_{ij} p_{ij} \text{ in } (p_{ij})_{1 \le i,j \le v} \\ &\text{subject to } \sum_{j=1}^{v} p_{ij} = 1 \ \forall i, \quad \sum_{i=1}^{v} p_{ij} = 1 \ \forall j, \quad p_{ij} \ge 0 \ \forall i, j. \end{aligned} \tag{3.2}$$

Here, p_{ij} corresponds to $v\mathbb{P}[X_1 = s_{1,i}, X_2 = s_{2,j}]$ if neither ϱ_1 nor ϱ_2 have multiple points at the same location. More generally, if we allow them to have multiple points, say $s_{1,i_1}, \dots, s_{1,i_r}$ are all the points of ϱ_1 that share some location $s \in E'$ and $s_{2,j_1}, \dots, s_{2,j_{\tilde{r}}}$ are all the points of ϱ_2 that share some location $\tilde{s} \in E'$, where $r, \tilde{r} \ge 1$, then $\sum_{k=1}^{r} \sum_{l=1}^{\tilde{r}} p_{i_k j_l}$ corresponds to $v\mathbb{P}[X_1 = s, X_2 = \tilde{s}]$, and it is of no importance how the mass is distributed among the $p_{i_k j_l}$, $1 \le k \le r$, $1 \le l \le \tilde{r}$, because the $d_{i_k j_l}$ are all equal.

The minimization in (3.2) is the assignment problem from linear programming with cost matrix $\frac{1}{v}(d_{ij})_{1 \le i,j \le v}$. It is a standard result that this problem always has a solution (p_{ij}) with $p_{ij} \in \{0, 1\}$ for all $i, j \in \{1, \dots, v\}$ (see Papadimitriou and Steiglitz (1982), Section 11.2 and the Corollary to Theorem 13.3). Hence the minimum in Problem (3.2) is equal to the minimum in the same problem with the additional constraint that $p_{ij} \in \{0, 1\}$ for all $i, j \in \{1, 2, \dots, v\}$, which can be written as

$$\min_{\pi \in \Sigma_v} \frac{1}{v} \sum_{i=1}^{v} d_{i,\pi(i)}$$

without any side conditions.

(iii) We first introduce some notation. For any $v \in \mathbb{N}$, denote by $\mathcal{P}_v$ the set of all subsets of $\{1, 2, \dots, v\}$, and for $s_1, \dots, s_v \in E'$, denote by $\mathcal{P}(s_1, \dots, s_v)$ the set of all subsets of $\{s_1, \dots, s_v\}$. For $\varrho_1, \varrho_2 \in \mathfrak{N}'$ with $|\varrho_1| = |\varrho_2| = v \ge 1$ and representations $\sum_{i=1}^{v} \delta_{s_{1,i}}$ and $\sum_{i=1}^{v} \delta_{s_{2,i}}$, it is easy to see that the Prohorov distance is simplified to

$$\begin{aligned} d_P(\varrho_1, \varrho_2) &= \inf\bigl\{\varepsilon > 0;\, \varrho_1(A) \le \varrho_2(A^\varepsilon) \text{ for all } A \in \mathcal{P}(s_{1,1}, \dots, s_{1,v})\bigr\} \\ &= \min\bigl\{\varepsilon \ge 0;\, \varrho_2\bigl(\textstyle\bigcup_{i \in I} \mathbb{B}(s_{1,i}, \varepsilon)\bigr) \ge |I| \text{ for all } I \in \mathcal{P}_v\bigr\}, \end{aligned} \tag{3.3}$$

using that $A^\varepsilon = E'$ for $\varepsilon > 1$ in the first equality.

We first show that $d_P(\varrho_1, \varrho_2) \le \min_\pi \max_i d_0(s_{1,i}, s_{2,\pi(i)})$ for any such ϱ_1 and ϱ_2. Write δ for the right hand side, and choose a minimizing $\pi \in \Sigma_v$, so that $d_0(s_{1,i}, s_{2,\pi(i)}) \le \delta$ for $1 \le i \le v$. It then follows immediately that $\varrho_2\bigl(\bigcup_{i \in I} \mathbb{B}(s_{1,i}, \delta)\bigr) \ge |I|$ for any $I \in \mathcal{P}_v$, because $\bigcup_{i \in I} \mathbb{B}(s_{1,i}, \delta)$ contains at least the points $s_{2,\pi(i)}$ of ϱ_2 with $i \in I$. Hence $d_P(\varrho_1, \varrho_2) \le \delta$ by Equation (3.3).

The converse inequality is a bit more difficult to show. For $\varrho_1, \varrho_2 \in \mathfrak{N}'$ with $|\varrho_1| = |\varrho_2| = v \geq 1$ and representations $\sum_{i=1}^{v} \delta_{s_{m,i}}$, $m = 1, 2$, and for $\varepsilon > 0$, we say that *Condition* $C_v(\varrho_1, \varrho_2; \varepsilon)$ *holds* if $\varrho_2\big(\bigcup_{i\in I} \mathbb{B}(s_{1,i}, \varepsilon)\big) \geq |I|$ for all $I \in \mathcal{P}_v$. We claim that Condition $C_v(\varrho_1, \varrho_2; \varepsilon)$ implies that we can pair the points of ϱ_1 with the points of ϱ_2 via a permutation $\pi \in \Sigma_v$ in such a way that $d_0(s_{1,i}, s_{2,\pi(i)}) \leq \varepsilon$ for all $i \in \{1, \ldots, v\}$. Setting in particular $\varepsilon := d_P(\varrho_1, \varrho_2)$ then yields that $\min_\pi \max_i d_0(s_{1,i}, s_{2,\pi(i)}) \leq d_P(\varrho_1, \varrho_2)$.

Our claim can be shown by induction over v. For $v = 1$ we have only two points $s_{1,1}$ and $s_{2,1}$ which have d_0-distance at most ε, so we pair them. Assume the statement is true for $v-1$ and show the statement for v: amongst all the sets $I \in \mathcal{P}_v$ that satisfy *equality* in Condition $C_v(\varrho_1, \varrho_2; \varepsilon)$ (note that equality always holds for $I = \{1, 2, \ldots, v\}$), choose one with minimal cardinality $r \in \{1, 2, \ldots, v\}$ and denote it by I_0. Choose an arbitrary $i_0 \in I_0$ and pair s_{1,i_0} with an arbitrary point s_{2,j_0} of ϱ_2 that lies in $\mathbb{B}(s_{1,i_0}, \varepsilon)$ (i.e. set $\pi(i_0) := j_0$). Remove then s_{1,i_0} and s_{2,j_0} from their respective point processes, that is consider the new point processes $\tilde{\varrho}_1 := \varrho_1 - \delta_{s_{1,i_0}}$ and $\tilde{\varrho}_2 := \varrho_2 - \delta_{s_{2,j_0}}$, and set $I_1 := I_0 \setminus \{i_0\}$. All that is left to do is to prove that Condition $C_{v-1}(\tilde{\varrho}_1, \tilde{\varrho}_2; \varepsilon)$ holds, because by the induction hypothesis we can then pair the $v-1$ points of $\tilde{\varrho}_1$ with those of $\tilde{\varrho}_2$ without exceeding the distance ε, which together with the pair (s_{1,i_0}, s_{2,j_0}) yields the required total pairing.

To show $C_{v-1}(\tilde{\varrho}_1, \tilde{\varrho}_2; \varepsilon)$, choose a subset I of $\{1, 2, \ldots, v\} \setminus \{i_0\}$. For $r = |I_0| = 1$, obviously $\tilde{\varrho}_2\big(\bigcup_{i\in I\cap I_1} \mathbb{B}(s_{1,i}, \varepsilon)\big) = 0 = |I \cap I_1|$. For $r \geq 2$, we have

$$\tilde{\varrho}_2\Big(\bigcup_{i\in I\cap I_1} \mathbb{B}(s_{1,i}, \varepsilon)\Big) \geq \varrho_2\Big(\bigcup_{i\in I\cap I_1} \mathbb{B}(s_{1,i}, \varepsilon)\Big) - 1 \geq |I \cap I_1|, \tag{3.4}$$

where the second inequality follows by the minimal cardinality of I_0 and $|I \cap I_1| < |I_0|$. Furthermore, we have for any $r \geq 1$,

$$\begin{aligned}\tilde{\varrho}_2\Big(\bigcup_{i\in I} \mathbb{B}(s_{1,i}, \varepsilon) \setminus \bigcup_{i\in I\cap I_1} \mathbb{B}(s_{1,i}, \varepsilon)\Big) \\ \geq \varrho_2\Big(\bigcup_{i\in I\cup I_0} \mathbb{B}(s_{1,i}, \varepsilon) \setminus \bigcup_{i\in I_0} \mathbb{B}(s_{1,i}, \varepsilon)\Big) \geq |I \cup I_0| - |I_0| = |I \setminus I_1|,\end{aligned} \tag{3.5}$$

where the second inequality holds, because $\varrho_2\big(\bigcup_{i\in I_0} \mathbb{B}(s_{1,i}, \varepsilon)\big) = |I_0|$. Adding of Inequalities (3.4) and (3.5) yields $\tilde{\varrho}_2\big(\bigcup_{i\in I} \mathbb{B}(s_{1,i}, \varepsilon)\big) \geq |I|$, and hence Condition $C_{v-1}(\tilde{\varrho}_1, \tilde{\varrho}_2; \varepsilon)$ is shown. □

The relationships between the metrics defined on $\mathfrak{N}'$ are summarized in Figure 3.1.1. They all follow directly either from the definitions ($d_{\widetilde{P}} \leq d_P$) or from Lemma 3.1.A ($d_{\widetilde{W}} \leq d_P$) or from Figure 2.4.1 (all other relationships). Many of the "missing arrows" could easily be filled with non-uniform bounds that include in their formulations the total number of points in either of the point measures (if the numbers are equal). Thus, for example, $d_P(\varrho_1, \varrho_2) \leq v d_{\widetilde{W}}(\varrho_1, \varrho_2)$, where $v = |\varrho_1| = |\varrho_2|$ if the point counts of ϱ_1 and ϱ_2 are the same, and $v = 1$ otherwise.

We next examine the topologies generated by the various metrics. The following convergence result can be obtained. Recall that $\varrho_n \to \varrho$ weakly if $\int f \, d\varrho_n \to \int f \, d\varrho$ for every bounded continuous function $f : E' \to \mathbb{R}$, and $\varrho_n \to \varrho$ vaguely if $\int f \, d\varrho_n \to \int f \, d\varrho$ for every continuous function $f : E' \to \mathbb{R}$ that has compact support. Obviously, the two

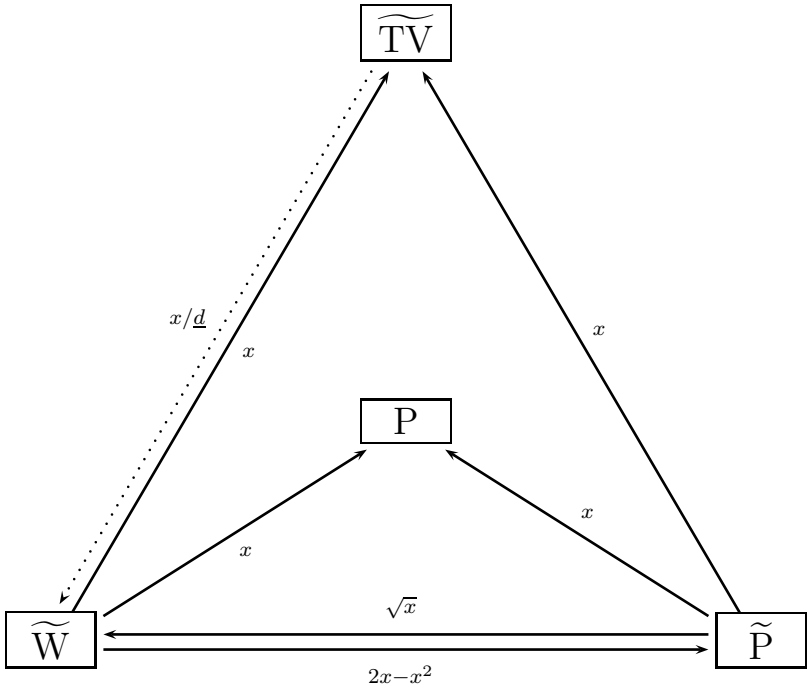

FIGURE 3.1.1: RELATIONSHIPS AMONG THE METRICS ON $\mathfrak{N}'$. For detailed explanations see Figure 2.4.1

concepts are equivalent on the compact space E', but they can of course also be defined in the same way on the more general space E for any finite measures.

Theorem 3.1.B. *Let $\varrho, \varrho_1, \varrho_2, \ldots \in \mathfrak{N}'$, and write $r_0 := \min\{d_0(s,t);\, \varrho(\{s\}) \geq 1, \varrho(\{t\}) \geq 1, s \neq t\}$. Then the following statements are equivalent.*

(i) $\varrho_n \longrightarrow \varrho$ *vaguely;*

(ii) $\varrho_n \longrightarrow \varrho$ *weakly;*

(iii) $\varrho_n(B) \to \varrho(B)$ *for all sets* $B \in \mathcal{B}'$ *with* $\varrho(\partial B) = 0$*;*

(iv) For every $\varepsilon \in (0, r_0/2)$ *there is an* $n_0 \in \mathbb{N}$ *such that, for* $n \geq n_0$, $|\varrho_n| = |\varrho|$ *and* $\varrho_n\big(\mathbb{B}(s,\varepsilon)\big) = \varrho(\{s\})$ *for every point* s *of* ϱ*;*

(v) $d_P(\varrho_n, \varrho) \longrightarrow 0$*;*

(vi) $d_{\widetilde{P}}(\varrho_n, \varrho) \longrightarrow 0$*;*

(vii) $d_1(\varrho_n, \varrho) \longrightarrow 0$.

Proof. The equivalence of (i), (ii) and (iii) is well-known for general finite measures. As mentioned above the equivalence of (i) and (ii) is obvious, and for the equivalence of (ii) and (iii) see for example Daley and Vere-Jones (1988), Theorem A.2.3.II.

(iii) $\Longrightarrow$ *(iv):* By (iii) we have in particular that $\varrho_n(E') \to \varrho(E')$. Hence there is an $N' \in \mathbb{N}$ such that $|\varrho_n| = |\varrho|$ for any $n \geq N'$. Let then $\varepsilon \in (0, r_0/2)$ and $s \in \operatorname{supp}(\varrho)$, that is $s \in E'$ with $\varrho(\{s\}) \geq 1$. By the choice of ε, we have $\varrho\big(\partial(\mathbb{B}(s,\varepsilon))\big) = 0$, so that, by (iii), there is an $N(s) \in \mathbb{N}$ such that $\varrho_n(\mathbb{B}(s,\varepsilon)) = \varrho(\mathbb{B}(s,\varepsilon)) = \varrho(\{s\})$ for any $n \geq N(s)$. Choosing $n_0 := N' \vee \max\{N(s); s \in \operatorname{supp}(\varrho)\}$ yields the result claimed.

(iv) $\Longrightarrow$ *(v):* We use the formula for the Prohorov metric that is given in Lemma 3.1.A.

Let $\varepsilon \in (0, r_0/2)$. By (iv), there is an n_0 such that, for every $n \geq n_0$, $|\varrho_n| = |\varrho|$ and for each location in E' that contains $k \geq 1$ points of ϱ there are exactly k points of ϱ_n within distance ε; these points are farther away than ε from any other location in the space that contains any points of ϱ. Hence we can pair the points of $\varrho_n = \sum_{i=1}^{v} \delta_{s_{n,i}}$ and $\varrho = \sum_{i=1}^{v} \delta_{s_i}$ for any $n \geq n_0$ in such a way that the maximal distance between the points in each pair is at most ε. Thus

$$d_P(\varrho_n, \varrho) = \min_{\pi \in \Sigma_v} \max_{1 \leq i \leq v} d_0(s_{n,i}, s_{\pi(i)}) \leq \varepsilon$$

for every $n \geq n_0$, which yields the statement.

(v) $\Longrightarrow$ (vi): Follows immediately by $d_{\widetilde{P}} \leq d_P$.

(vi) $\Longrightarrow$ (vii): Follows immediately by $d_1 \leq 2 d_{\widetilde{P}}$.

(vii) $\Longrightarrow$ (ii): Follows with a slight adaptation from Theorem 2.4.D. The convergence $d_1(\varrho_n, \varrho) \to 0$ implies that there is an $n_0 \in \mathbb{N}$ such that $|\varrho_n| = |\varrho|$ for any $n \geq n_0$. For $\varrho = 0$, this implies (ii) directly. For $|\varrho| \geq 1$, it implies that $d_{BW}(\varrho_n/|\varrho|, \varrho/|\varrho|) \to 0$ for $n \to \infty$, where we only consider $n \geq n_0$, so that the d_{BW}-term is always well-defined as a bounded Wasserstein distance between two probability measures. Theorem 2.4.D now yields that $\varrho_n/|\varrho| \xrightarrow{w} \varrho/|\varrho|$, and hence also $\varrho_n \xrightarrow{w} \varrho$. □

Remark 3.1.C. It is easily checked that the equivalence of the statements (ii) to (vii) is still true if we consider finite point measures on the space E (and replace $\mathcal{B}'$ by $\mathcal{B}$ in statement (iii))

In summary, Theorem 3.1.B and Remark 3.1.C yield for the metrics d_P, $d_{\widetilde{P}}$, and d_1 that they are all equivalent, and metrize the weak topology on $\mathfrak{N}'$. Since E' is compact, the latter is the same as the vague topology, which is our canonical topology on $\mathfrak{N}'$ and in this sense the "right topology to use" (cf. Section 2.2).

We drop the relative total variation metric as it is in fact contained in the definition of the d_1-metric. If we ignore the metric given on E' and consider the discrete metric instead, the corresponding d_1-metric is just the relative total variation metric. Furthermore we drop the relative Prohorov metric, as it follows the same idea as the Prohorov metric, but does not have the nice representation given in Lemma 3.1.A.

The relative Prohorov metric might sometimes be interesting for correcting the following "misbehavior" of the Prohorov metric when many points are involved. Consider $E' := [0, 2]$, equiped with the Euclidean metric trimmed at 1 as the metric d_0. Let $(\varrho_n)_n$ and $(\sigma_n)_n$ be two sequences of finite point measures on E' that are given by $\varrho_n := \sum_{i=1}^{n} \delta_{i/n}$ and $\sigma_n := \sum_{i=1}^{n-1} \delta_{i/n} + \delta_2$. We then have $d_1(\varrho_n, \sigma_n) = 1/n$ (see the proof of Lemma 3.3.B) and $d_{\widetilde{P}}(\varrho_n, \sigma_n) = 1/n$ (directly by the definition), but obviously, by Lemma 3.1.A, $d_P(\varrho_n, \sigma_n) = 1$ for every n. So in a sense, the Prohorov metric is sensitive to "outliers", which may sometimes be undesirable.

Of the two metrics left over (d_1 and d_P), the metric of our choice will later on always be the d_1-metric, as it was mentioned before. This is for several reasons, which are detailed in the next section, where we compare the Wasserstein metrics built on d_1 and on d_P with one another. Nevertheless, the Prohorov metric is a very important and useful metric, too. It is the more traditional metric on $\mathfrak{N}'$ and often appears in textbooks on point processes, such as Kallenberg (1986) and Daley and Vere Jones (1988), usually without

the somewhat artificial trimming at distance 1. Also it lacks, because of Lemma 3.1.A, the usual shortcoming of a Prohorov metric of being difficult to handle.

We finish this section by giving some of the topological properties of the space $(\mathfrak{N}', d_1)$. It is easily seen that the same properties hold also for $(\mathfrak{N}', d_P)$.

Proposition 3.1.D. *The metric d_1 generates the vague topology on $\mathfrak{N}'$. $(\mathfrak{N}', d_1)$ is a complete separable metric space that is locally compact (and hence σ-compact).*

Proof. The first statement follows from Theorem 3.1.B and the fact that the vague topology is completely described by the vague convergence of sequences in $\mathfrak{N}'$. The latter statement is true by Theorem X.6.2 in Dugundji (1966), because $(\mathfrak{N}', \mathcal{T}_{\mathfrak{N}'})$ satisfies the first axiom of countability, which is a weak consequence of Proposition A.2.C.

Also by Proposition A.2.C, $\mathfrak{N}'$ equipped with the vague topology is separable.

By the compactness of E', it can be seen that $C_v := \{\varrho \in \mathfrak{N}';\ \varrho(E') = v\}$ is compact for every $v \in \mathbb{Z}_+$. Furthermore the C_v are obviously open, so that any $\varrho \in C_v$ has C_v as a compact neighborhood. The σ-compactness follows either by writing $\mathfrak{N}'$ as the union of the C_v or, as in the proof of Proposition A.2.B, from the local compactness, using that every separable metric space is Lindelöf (see Dugundji (1966), Theorem IX.5.6).

The completeness is the only one of the properties mentioned that depends on the metric. The C_v defined above are complete (in any metric), since they are compact. But any d_1-Cauchy sequence in $\mathfrak{N}'$ ultimately ends up in one of the C_v (i.e. there is a $v \in \mathbb{Z}_+$ and an $n_0 \in \mathbb{N}$ such that $\varrho_n \in C_v$ for all $n \geq n_0$), hence it has a tail sequence that is a Cauchy sequence in C_v, and therefore converges. □

3.1.2 A note on computing distances between point measures

For two point processes ϱ_1 and ϱ_2 with v points each, d_1 and d_P are both defined by minimization over a set of cardinality $v!$, so at first glance, they both involve computationally hard problems. However, by the so-called Hungarian method from linear programming, the d_1-distance can be calculated in $O(v^3)$ steps only, and a basic implementation of the method programmed by the author as a java applet (see the screenshot in Figure 3.1.2) shows that distances for a moderate number of points (e.g. up to 100 in each point measure) can be calculated on a modern PC within fractions of a second. The Hungarian method is essentially an adaptation of the simplex algorithm to the assignment situation at hand; for a detailed account see Papadimitriou and Steiglitz (1982), Section 11.2.

For the d_P-metric, the situation seems not to be so friendly. The problem can also be formulated as a linear program, but a more complicated one, especially as a true integer (more precisely, binary) program has to be solved, whereas for the d_1-metric we can solve a simple linear program, and there automatically turns out to be an integer solution (compare the proof of Lemma 3.1.A(ii)). There does not seem to be any such possibility of simplification for d_P. Note that the general binary linear programming problem is NP-complete; see the proof of the corollary to Theorem 15.1 in Papadimitriou and Steiglitz (1982).

3.1.3 A note on defining metrics on $\mathfrak{N}$

It may seem a bit unaesthetic that we do not define metrics between point process distributions on the more general space E, but there does not seem to be a simple way that is

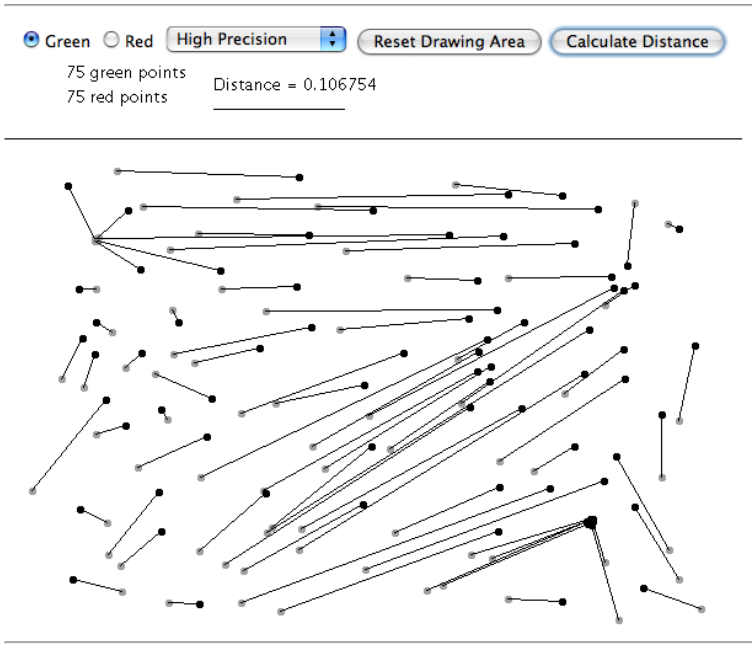

FIGURE 3.1.2: SCREENSHOT FROM DEONEAPPLET. The lower part displays the optimal pairing for the selected point patterns with regard to the d_1-distance; in the upper part some additional information is given. Note that the grey-scale values of the points have been altered for better contrast between what in the original are green points (now light-grey) and red points (now black). The applet and its source code can be obtained from http://www.math.unizh.ch/∼schumi/Java/DeOneApplet.html.

satisfying for our purposes. In this subsection, we briefly present some ideas for extending the metrics defined on $\mathfrak{N}'$, and explain in what respects they are more or less suitable.

First of all, one might look for direct ways of extending the metrics from the previous section to $\mathfrak{N}$. For the d_1-metric, it is not clear how to do so. A naive definition is, for $\varrho_1, \varrho_2 \in \mathfrak{N}$ with $|\varrho_1| = |\varrho_2| = \infty$ and representations $\sum_{i=1}^{\infty} \delta_{s_{1,i}}$ and $\sum_{i=1}^{\infty} \delta_{s_{2,i}}$, respectively, to set

$$\overline{d}_1(\varrho_1, \varrho_2) := \inf_{\pi \in \Sigma} \lim_{v \to \infty} \frac{1}{v} \sum_{i=1}^{v} d_0(s_{1,i}, s_{2,\pi(i)}),$$

where Σ is the set of all permutations on $\mathbb{N}$ for which the above limit exists. For other $\varrho_1, \varrho_2 \in \mathfrak{N}$, $\overline{d}_1(\varrho_1, \varrho_2)$ is just defined in the "natural" way, that is, it is set equal to 1 if $|\varrho_1|$ and $|\varrho_2|$ do not agree, and it is defined in the same way as $d_1(\varrho_1, \varrho_2)$ if $|\varrho_1|$ and $|\varrho_2|$ are finite.

Trying to define $\overline{d}_1$ in such a way is, of course, a very bad idea, because it is not even a metric (e.g. there are many different point measures with "distance" 0), the definition depends on the enumeration of the points, and it fails to capture the concept of vague convergence on $\mathfrak{N}$. This last point is seen by letting for example on $E := \mathbb{R}^2$, equipped with the truncated Euclidean metric, $\varrho_n := \sum_{i=1}^{n} \delta_{s_i} + \sum_{i=n+1}^{\infty} \delta_{t_i}$ and $\varrho := \sum_{i=1}^{\infty} \delta_{s_i}$, where $s_i := (0, i)$ and $t_i := (1, i)$ for $i \in \mathbb{N}$. We then have $\varrho_n \to \varrho$ vaguely, but $\overline{d}_1(\varrho_n, \varrho) = 1$ for all n. There are similar ways to extend the d_1-metric "naively" which all have at least some of the problems mentioned above.

For the d_P-metric, on the other hand, the corresponding extension seems to be a little less problematic from a theoretical point of view, but it is, of course, even stronger than $\overline{d}_1$. For $\varrho_1, \varrho_2 \in \mathfrak{N}$ with $|\varrho_1| = |\varrho_2| = \infty$ and representations $\sum_{i=1}^{\infty} \delta_{s_{1,i}}$ and $\sum_{i=1}^{\infty} \delta_{s_{2,i}}$, respectively, set

$$\overline{d}_P(\varrho_1, \varrho_2) := \inf_{\pi \in \Sigma} \max_{i \in \mathbb{N}} d_0(s_{1,i}, s_{2,\pi(i)}),$$

where Σ is the set of all permutations on $\mathbb{N}$. As above, we define $\overline{d}_P(\varrho_1, \varrho_2)$ for other $\varrho_1, \varrho_2 \in \mathfrak{N}$ in the natural way. Again the example from above applies to show that $\overline{d}_P$ is too strong. We obviously here have $d_P(\varrho_n, \varrho) = 1$.

It then becomes clear that in order to obtain a metric that metrizes vague convergence on $\mathfrak{N}$, we have to associate weights that go to zero to the different point pairs. It is unwise to do so directly by weighting the point pairs according to their number in some enumeration of the pairs (this gives no metric in general, and the resulting function will depend on the enumeration). Thus, we associate weights, essentially, according to the position of the points in E. The most convenient way to do so is given by the construction proposed in Daley and Vere-Jones (1988), Section A.2.6, which we slightly adapt to our special situation (point measures and metrics that are bounded by 1). Note that Daley and Vere-Jones applied this construction only for extending the Prohorov metric, with good reason, as we will see below.

Assume that we have a metric $d \leq 1$ between finite point measures on arbitrary *metrically* bounded subsets of E. We extend this to a metric $\overline{d} \leq 1$ on $\mathfrak{N}$ by choosing an arbitrary $s_0 \in E$ and setting

$$\overline{d}(\varrho_1, \varrho_2) := \int_0^{\infty} d\big(\varrho_1|_{\mathbb{B}(s_0,r)}, \varrho_2|_{\mathbb{B}(s_0,r)}\big) e^{-r}\, dr, \tag{3.6}$$

for any $\varrho_1, \varrho_2 \in \mathfrak{N}$ that have finite total mass on every ball $\mathbb{B}(s_0, r)$ (note that for a general lcscH E the closed balls need not be compact). It is easy to see that $\overline{d}$ is a metric on this smaller space $\mathfrak{N}^-$ of point measures that is furthermore bounded by 1. However, this restriction to $\mathfrak{N}^-$, as well as the fact that the metric defined (although not the topology it generates) depends on the point s_0, for which there is in general no natural choice, makes this definition very awkward.

If we consider the same definition on $E = \mathbb{R}^D$ and take the balls with respect to an (untrimmed) metric that is induced by a norm on $\mathbb{R}^D$, the situation looks better, since then the closed balls are compact and we have the natural choice $s_0 := 0$ for their centers. Under these circumstances, $\overline{d}$ is indeed defined on all of $\mathfrak{N}$ and generates the vague topology. So from a topological point of view this is interesting. Unfortunately, for measuring concrete distances this metric is very unintuitive and rather difficult to handle. Also, it provides an extension of d with a strong Prohorov flavor, in that a few not so small distances between points make the whole $\overline{d}$-distance large. This means that nice properties which we obtain for the metric $d = d_1$, say, (that are not shared by d_P) are usually destroyed by this extension. Consider for example the point measures $\varrho_1 := \sum_{i=1}^{n} \delta_{i/n}$ and $\varrho_2 := \sum_{i=1}^{n-1} \delta_{i/n} + \delta_2$ on $E = \mathbb{R}$. They satisfy $d_1(\varrho_1, \varrho_2) = 1/n$, but if we compute $\overline{d}$ based on d_1, we obtain $\overline{d}(\varrho_1, \varrho_2) = \int_1^2 e^{-r}\, dr + \int_2^{\infty} \frac{1}{n} e^{-r}\, dr = e^{-1} - \frac{n-1}{n} e^{-2}$, which does not become small for n large.

Thus, the construction in Equation (3.6) is most natural for extending the Prohorov metric between point measures on $\mathbb{R}^D$, but even in this case, the extended metric $\overline{d}$ is

not very nice in view of the distances it returns. First, it is not an extension in a strict sense, because for finite point measures, it does in general not agree with the Prohorov distance. Secondly, the distances measured in $\overline{d}$ are usually difficult to interprete because of all the balls $\mathbb{B}(0,r)$ that have been stacked on one another for its definition. Finally, $\overline{d}$ is not invariant with respect to simultaneous translations of the point measures, which would often be desirable.

3.2 Definition and elementary properties of the Barbour-Brown metric

We define the Barbour-Brown metric on $\mathfrak{P}(\mathfrak{N}')$ as the Wasserstein metric based on the metric d_1 on $\mathfrak{N}'$.

Definition. The *Barbour-Brown metric* d_2 (or short *d_2-metric*) on $\mathfrak{P}(\mathfrak{N}')$ with respect to d_0 on E' is defined by

$$d_2(P,Q) := \sup_{f\in\mathfrak{F}_2} \left| \int f\, dP - \int f\, dQ \right|$$

for any probability measures P and Q on $\mathfrak{N}'$, where

$$\mathfrak{F}_2 := \big\{ f : \mathfrak{N}' \to \mathbb{R}\,;\ |f(\varrho_1) - f(\varrho_2)| \le d_1(\varrho_1,\varrho_2) \text{ for all } \varrho_1,\varrho_2 \in \mathfrak{N}' \big\}.$$

The following proposition summarizes some results in connection with the Barbour-Brown metric.

Theorem 3.2.A.

(i) For probability measures P and Q on $\mathfrak{N}'$, we have that

$$d_2(P,Q) = \min_{\substack{\xi_1\sim P\\ \xi_2\sim Q}} \mathbb{E}\, d_1(\xi_1,\xi_2).$$

(ii) The Barbour-Brown metric metrizes the weak convergence on $\mathfrak{P}(\mathfrak{N}')$, that is, for point processes $\xi,\xi_1,\xi_2,\ldots$ on E', we have that $\xi_n \xrightarrow{\mathscr{D}} \xi$ iff $d_2\big(\mathscr{L}(\xi_n),\mathscr{L}(\xi)\big) \longrightarrow 0$.

Proof. (i) follows from Theorem 2.4.B. The minimum exists, because $\mathfrak{N}'$ is complete by Proposition 3.1.D.

(ii) follows from Theorem 2.4.D, using that $d_1 \le 1$ and that hence d_2 is the bounded Wasserstein metric with respect to d_1. Also used is the fact from Proposition 3.1.D that d_1 metrizes the vague topology on $\mathfrak{N}'$, that is, the topology used to define weak convergence in $\mathfrak{P}(\mathfrak{N}')$ (see Subsection 2.2.3). □

The first statement of the theorem provides a basic way of bounding the d_2-distance. If there is a natural coupling of the distributions P and Q, that is if there is a pair (ξ,η) of point processes with $\xi \sim P$, $\eta \sim Q$, and with a common distribution that brings realizations of ξ and η close together in terms of the d_1-metric, then a promising first estimate for the d_2-distance between P and Q is given by

$$d_2(P,Q) \le \mathbb{P}[|\xi| \ne |\eta|] + \mathbb{E}\big(d_1(\xi,\eta) 1_{\{|\xi|=|\eta|\ge 1\}}\big). \tag{3.7}$$

The second statement of Theorem 3.2.A enables us to prove convergence statements by calculating d_2-upper bounds that go to zero.

Towards the end of Section 3.1, the possibility was left open to choose the Prohorov metric on $\mathfrak{N}'$ instead of the d_1-metric. We define then in the analogous way the metric d_2^* as the Wasserstein metric based on d_P.

Definition. The metric d_2^* on $\mathfrak{P}(\mathfrak{N}')$ with respect to d_0 on E' is defined by

$$d_2^*(P,Q) := \sup_{f\in\mathfrak{F}_2^*} \left| \int f\, dP - \int f\, dQ \right|$$

for any probability measures P and Q on $\mathfrak{N}'$, where

$$\mathfrak{F}_2^* := \bigl\{ f : \mathfrak{N}' \to \mathbb{R}\,;\ |f(\varrho_1) - f(\varrho_2)| \le d_P(\varrho_1,\varrho_2) \text{ for all } \varrho_1,\varrho_2 \in \mathfrak{N}' \bigr\}.$$

The same results as in Theorem 3.2.A hold also for the metric d_2^*, with the obvious replacement of d_1 by d_P in statement (i) (since $d_1 \le d_P$ and the two metrics are equivalent, the completeness of $(\mathfrak{N}', d_1)$ implies the completeness of $(\mathfrak{N}', d_P)$, so that the minimum exists).

As stated farther above, the d_2-metric is in later chapters the only metric on $\mathfrak{P}(\mathfrak{N}')$ that we consider. However, the d_2^*-metric may sometimes prove more useful in applications. So it is well worth spending a few thoughts on this metric as well.

Note, first of all, that the two metrics have a lot in common. They are both metrics on $\mathfrak{N}'$ that are bounded by 1, return ultimately to the metric d_0 on E', and metrize the weak topology on $\mathfrak{P}(\mathfrak{N}')$, so that in particular, they are topologically equivalent.

Based on the metrics on $\mathfrak{N}'$ from which they are derived, they measure distances a bit differently. In fact, $d_2 \le d_2^*$ (because obviously $\mathfrak{F}_2 \subset \mathfrak{F}_2^*$), which means that we usually get better upper bounds for d_2, but that the upper bounds for d_2^* can be used as upper bounds for terms of the form $|P(f) - Q(f)|$ for a bigger class of functions f.

In what follows, the relevance of both of these points is briefly illustrated. Let the space $E' = \{\alpha_i;\, 1 \le i \le n\}$ be a finite set and d_0 the discrete metric. Let $q \in (0,1)$, and consider point processes $\xi_n := \sum_{i=1}^n X_i \delta_{\alpha_i}$ and $\eta_n := \sum_{i=1}^n Y_i \delta_{\alpha_i}$ on E', where the X_i are independent indicators with expectation q and the Y_i are independent $\mathrm{Po}(q)$-distributed random variables. A process of the form of ξ_n is usually referred to as a Bernoulli process on E' with parameter q, while η_n is simply a $\mathrm{Po}\bigl(q\sum_{i=1}^n \delta_{\alpha_i}\bigr)$-process. Theorem 4.3.C in the next chapter yields that $d_2(\mathscr{L}(\xi_n), \mathscr{L}(\eta_n)) = O(q \wedge nq^2)$ (and Theorem 3.D.1 in Barbour, Holst, and Janson (1992) in connection with Inequality (3.15) implies that the order for the upper bound is sharp). However, for d_2^* it is not possible to recover the "magic factor" $1/nq$ (for the terminology, see Remark 4.4.B): all that we can get is $d_2^*(\mathscr{L}(\xi_n), \mathscr{L}(\eta_n)) = O(nq^2)$. This is seen by realizing that d_P on $\mathfrak{N}'$ with respect to the

discrete metric on E' is the discrete metric on $\mathfrak{N}'$. Hence

$$\begin{aligned} d_2^*\big(\mathscr{L}(\xi_n),\mathscr{L}(\eta_n)\big) &= d_{TV}\big(\mathscr{L}(\xi_n),\mathscr{L}(\eta_n)\big) \\ &= \min_{\substack{X_i\sim\mathrm{Be}(q),\perp\!\!\!\perp \\ Y_i\sim\mathrm{Po}(q),\perp\!\!\!\perp}} \Big(1-\mathbb{P}\big[X_i=Y_i \text{ for all } 1\le i\le n\big]\Big) \\ &= \min_{\substack{X_i\sim\mathrm{Be}(q),\perp\!\!\!\perp \\ Y_i\sim\mathrm{Po}(q),\perp\!\!\!\perp}} \Big(1-\prod_{i=1}^n\big(1-\mathbb{P}[X_i\neq Y_i]\big)\Big) \\ &= 1-\Big(1-d_{TV}\big(\mathscr{L}(X_1),\mathscr{L}(Y_1)\big)\Big)^n \\ &= 1-\big(1-q(1-e^{-q})\big)^n \\ &\asymp nq^2, \end{aligned}$$

where the second and the fourth equality follow from Theorem 2.4.A, and the fifth equality is seen by

$$d_{TV}\big(\mathscr{L}(X_1),\mathscr{L}(Y_1)\big)=\frac{1}{2}\sum_{k=0}^{\infty}\Big|\mathbb{P}[X_1=k]-\mathbb{P}[Y_1=k]\Big|=\mathbb{P}[X_1=1]-\mathbb{P}[Y_1=1],$$

which follows from $\mathbb{P}[X_1=i]-\mathbb{P}[Y_1=i]<0$ for $i\neq 1$. Thus, d_2^* behaves here in exactly the same way as the total variation metric. The magic factor from Stein's method is lost.

We turn now to the illustration of the second point, namely that there are interesting functions f for which $|P(f)-Q(f)|$ is bounded by $d_2^*(P,Q)$, but not by $d_2(P,Q)$. Consider the function $f:\mathfrak{N}'\to\mathbb{R}$ which maps a finite point pattern on its diameter (or its maximal interpoint distance), that is $f(\varrho):=\max_{i,j} d_0(s_i,s_j)$ for any $\varrho=\sum_{i=1}^v \delta_{s_i}\in\mathfrak{N}'$. It is easily checked that $\frac{1}{2}f\in\mathfrak{F}_2^*$ (compare also Example 3 in Subsection 3.3.1), so that the d_2^*-bound above yields that

$$\big|\mathbb{E}f(\xi_n)-\mathbb{E}f(\eta_n)\big|=O(nq^2),$$

but no such statement is obtained from $d_2(\mathscr{L}(\xi_n),\mathscr{L}(\eta_n))=O(q\wedge nq^2)$, because f is not Lipschitz continuous with respect to d_1.

Since we do not calculate explicit bounds for d_2^*-distances later on, we finish this section by providing some tools for bounding $|P(f)-Q(f)|$ for functions $f\in\mathfrak{F}_2^*\setminus\mathfrak{F}_2$ based on upper bounds for $d_2(P,Q)$. The first result yields that many upper bounds obtained by the Stein local approach for the d_2-metric carry over to the case of the d_2^*-metric, except for the fact that any magic factors are lost, which is in accordance with what was obtained above.

Proposition 3.2.B. *The upper bounds for the d_2-distances of the form $d_2(\mathscr{L}(\xi),\mathrm{Po}(\lambda))$ obtained in Theorem 4.3.A, Theorem 5.2.A, Theorem 6.3.B, and Theorem 7.2.A hold also for the corresponding d_2^*-distances $d_2^*(\mathscr{L}(\xi),\mathrm{Po}(\lambda))$, provided that any of the magic factors $M_1(\lambda)$ and $M_2(\lambda)$ (see Section 4.2) are replaced by* 1.

Proof. We give only the instructions how the corresponding proofs have to be altered.

As far as Theorem 4.3.A is concerned, the special properties of the functions $f\in\mathfrak{F}_2$ are only used in the estimations of $\|\Delta_1 h\|_\infty$ and $\|\Delta_2 h\|_\infty$, where $h=h_f$ is the solution (4.6) to the Stein equation (4.5) (see Section 4.2 for the context). The upper bounds

by 1 for $f \in \mathfrak{F}_2^*$ are very easily obtained by adapting the proofs of the corresponding Inequalities (4.7) and (4.8).

In the proofs of Theorems 5.2.A and 6.3.B, we use Theorem 4.3.A and also a discretization argument. However, the latter is carried out in such a way that we actually estimate the d_2^*-distance and not the weaker d_2-distance.

Finally, for Theorem 7.2.A, a very similar argument holds as for Theorem 4.3.A. Special properties of $f \in \mathfrak{F}_2$ are only used for estimating $\|\Delta_1 h\|_\infty$ and $\|\Delta_2 h\|_\infty$, as well as for computing the Lipschitz constant C in Inequality (7.3). Obtaining the upper bound 1 if $f \in \mathfrak{F}_2^*$ instead of the magic factors if $f \in \mathfrak{F}_2$ is very easy for each of these terms. □

The following result provides a d_2-upper bound for a general d_2^*-distance term. Of particular interest is the case where at least one of the point processes is a Poisson process, which is almost the only case we are faced with in this thesis. It follows then that the d_2^*-term can be bounded by a term that is only slightly worse than the bound suggested in the special situation of Proposition 3.2.B. We briefly illustrate this point after the proof.

Proposition 3.2.C. *Let ξ and η be point processes on E'. Then*

$$d_2\big(\mathscr{L}(\xi),\mathscr{L}(\eta)\big) \le d_2^*\big(\mathscr{L}(\xi),\mathscr{L}(\eta)\big) \le m d_2\big(\mathscr{L}(\xi),\mathscr{L}(\eta)\big) + \mathbb{E}\big(|\eta| 1_{\{|\eta|>m\}}\big) \tag{3.8}$$

for any $m \in \mathbb{N}$. If η is a Poisson process with (finite) intensity measure λ, then

$$d_2^*\big(\mathscr{L}(\xi),\mathscr{L}(\eta)\big) \le m d_2\big(\mathscr{L}(\xi),\mathscr{L}(\eta)\big) + \frac{1}{3\sqrt{\pi m}} e^{-m} \tag{3.9}$$

for any $m \in \mathbb{N}$ with $m \ge e^2|\lambda|$.

Proof. The first inequality is immediate from the fact that $\mathfrak{F}_2 \subset \mathfrak{F}_2^*$. The second inequality can be shown as follows. Let $\tilde\xi$ and $\tilde\eta$ be arbitrary point processes, defined on the same probability space, with $\tilde\xi \stackrel{\mathscr{D}}{=} \xi$ and $\tilde\eta \stackrel{\mathscr{D}}{=} \eta$. We obtain

$$\mathbb{E} d_P(\tilde\xi,\tilde\eta) \le \mathbb{E}\big((|\tilde\eta| \vee 1) d_1(\tilde\xi,\tilde\eta)\big) \le m\mathbb{E}\big(d_1(\tilde\xi,\tilde\eta)\big) + \mathbb{E}\big(|\tilde\eta| 1_{\{|\tilde\eta|>m\}}\big)$$

for any $m \in \mathbb{N}$. Taking the minimum over all $\tilde\xi$ and $\tilde\eta$ with $\tilde\xi \stackrel{\mathscr{D}}{=} \xi$ and $\tilde\eta \stackrel{\mathscr{D}}{=} \eta$ yields the second inequality in (3.8). If η is a Po(λ)-process the expectation in (3.8) can be estimated by Lemma A.1.A as

$$\begin{aligned}
\mathbb{E}\big(|\eta| 1_{\{|\eta|>m\}}\big) &= |\lambda|\,\mathbb{P}[|\eta| \ge m] \\
&\le \frac{|\lambda|(m+1)}{\sqrt{2\pi m}(m+1-|\lambda|)} \left(\frac{|\lambda|}{m}\right)^m e^{m-|\lambda|} \\
&\le \frac{1}{3\sqrt{\pi m}} e^{-m}
\end{aligned} \tag{3.10}$$

for $m \in \mathbb{N}$ with $m \ge e^2|\lambda|$, where we used that $xe^{-x} \le e^{-1}$ for all $x \ge 0$ for the last inequality. □

For the sake of illustration assume that two sequences $(\xi_n)_n$ and $(\eta_n)_n$ of point processes are given, where η_n is a Poisson process with intensity measure λ_n. Let $r \in \mathbb{N}$, and set

$m := m_n := \lceil r\log(n) \vee e^2|\lambda_n| \rceil$ for every $n \in \mathbb{N}$, where $\lceil x \rceil := \min\{z \in \mathbb{Z}; z \geq x\}$ for $x \in \mathbb{R}$.

If $|\lambda_n| = O(\log(n))$, especially if $|\lambda_n|$ is bounded, Proposition 3.2.C yields that

$$d_2^*\big(\mathscr{L}(\xi_n), \mathscr{L}(\eta_n)\big) = O\bigg(\log(n) d_2\big(\mathscr{L}(\xi_n), \mathscr{L}(\eta_n)\big) + \frac{1}{n^r}\bigg) \quad \text{for } n \to \infty.$$

In most applications the term $1/n^r$ is negligible if $r \in \mathbb{N}$ is chosen large enough, so that the order of the d_2^*-distance is only worse by a logarithmic factor than the d_2-distance.

If, on the other hand, $|\lambda_n| = \omega(\log(n))$, that is $|\lambda_n|/\log(n) \to \infty$ for $n \to \infty$, then

$$d_2^*\big(\mathscr{L}(\xi_n), \mathscr{L}(\eta_n)\big) = O\Big(|\lambda_n| d_2\big(\mathscr{L}(\xi_n), \mathscr{L}(\eta_n)\big) + e^{-|\lambda_n|}\Big) \quad \text{for } n \to \infty.$$

Again, even more so than before, the second summand is usually negligible. So this time the order of the d_2^*-distance is worse by a factor $|\lambda_n|$ than the d_2-distance, which up to a remaining factor $\log^+(|\lambda_n|/2)$ corresponds to the removal of a magic factor $M_2(\lambda)$ from the whole estimate (note however that any factor $M_1(\lambda)$ we might have in the situation of Proposition 3.2.B is overcompensated here).

In summary, Inequality (3.9) of Proposition 3.2.C relates the d_2^*-distance between an arbitrary point process and a Poisson process to the d_2-distance between these two processes in much the same spirit as Proposition 3.2.B does their upper bounds obtained by Stein's method in special situations. However, for the cases in Proposition 3.2.B, the factor $M_1(\lambda_n)$ is replaced by a $O\big(1 \vee \sqrt{|\lambda_n|}\big)$-term instead of 1 and a (usually unimportant) logarithmic factor is introduced.

3.3 Consequences of d_2 upper bounds

As mentioned in the last section, a d_2-distance $d_2(\mathscr{L}(\xi_n), \mathscr{L}(\eta))$ that converges to zero characterizes the convergence $\xi_n \xrightarrow{\mathscr{D}} \eta$. However, as we have advocated in Chapter 1, a convergence statement is only a very weak result compared to a concrete distance estimate. We therefore examine in this section what other benefits we can gain from d_2-estimates such as the ones derived in Chapters 5 to 7.

In spite of the nice representations of the d_1- and d_2-metrics in Lemma 3.1.A and Theorem 3.2.A, the d_2-distance between two point process distributions seems to be rather difficult to grasp intuitively. Our main goal is therefore to bound simpler expressions by d_2-distances, so that any upper bound for the d_2-distance yields immediately an upper bound for the simpler expression. In this regard, the following lemma is important.

Lemma 3.3.A. *Let $f \in \mathfrak{F}_2$. Then*

$$d_W\big(\mathscr{L}(f(\xi)), \mathscr{L}(f(\eta))\big) \leq d_2\big(\mathscr{L}(\xi), \mathscr{L}(\eta)\big).$$

Proof. Let $h \in \mathfrak{F}_W := \big\{\tilde{h} : \mathbb{R} \to \mathbb{R};\ |\tilde{h}(x) - \tilde{h}(y)| \leq |x - y|\big\}$. Then obviously $h \circ f \in \mathfrak{F}_2$, whence

$$\big|\mathbb{E}h(f(\xi)) - \mathbb{E}h(f(\eta))\big| \leq d_2\big(\mathscr{L}(\xi), \mathscr{L}(\eta)\big),$$

which yields the statement. □

Lemma 3.3.A enables us to estimate distances between the distributions of certain point process statistics f directly by the corresponding d_2-distance between the underlying point process distributions. Note that the Wasserstein distance between distributions of $\mathbb{R}$-valued random variables X and Y has the nice intuitive representation as the area between the distribution functions of X and Y, that is

$$d_W\big(\mathscr{L}(X),\mathscr{L}(Y)\big) = \int_{-\infty}^{\infty} \big|\mathbb{P}[X \le x] - \mathbb{P}[Y \le x]\big|\, dx \tag{3.11}$$

(see Dudley (1989), Section 11.8, Problem 1, with hints).

We thus have reduced the distance $d_2\big(\mathscr{L}(\xi),\mathscr{L}(\eta)\big)$ to the much more intuitive distances $d_W\big(\mathscr{L}(f(\xi)),\mathscr{L}(f(\eta))\big)$, where $f \in \mathfrak{F}_2$. Our main concern is then that the class $\mathfrak{F}_2$ of admissible statistics contains enough interesting functions, which is demonstrated in Subsection 3.3.1.

Of course, there are other interesting quantities that are in some way functions of the closeness of two point process distributions, but cannot be written in the above way via $\mathfrak{F}_2$-functions. Subsection 3.3.2 deals with an example of such a quantity.

3.3.1 $\mathfrak{F}_2$-functions

In this subsection, a list of concrete $\mathfrak{F}_2$-functions is presented (see Table 3.3.1). The aim in compiling such a list is to give an idea of the richness of $\mathfrak{F}_2$ in view of statistically interesting functions. The list is by no means supposed to be representative for all of $\mathfrak{F}_2$, let alone complete.

The following simple lemma helps in establishing that a function is in $\mathfrak{F}_2$.

Lemma 3.3.B. *Let $\varrho_1 = \sum_{i=1}^v \delta_{s_i}$ and $\varrho_2 = \sum_{i=1}^v \delta_{t_i}$ be two point measures, and let a minimizing pairing of their points be given by $\pi_0 \in \Sigma_v$; that is, $\frac{1}{v}\sum_{i=1}^v d_0(s_i, t_{\pi_0(i)}) = d_1(\varrho_1,\varrho_2)$. Furthermore, let $\varrho^{(j)} := \sum_{i=1}^j \delta_{t_{\pi_0(i)}} + \sum_{i=j+1}^v \delta_{s_i}$ for $0 \le j \le v$, so that $\varrho^{(0)} = \varrho_1$ and $\varrho^{(v)} = \varrho_2$. Then*

$$d_1(\varrho_1,\varrho_2) = \sum_{j=1}^{v} d_1(\varrho^{(j-1)},\varrho^{(j)}).$$

Proof. Since $d_1(\varrho_1,\varrho_2) = \frac{1}{v}\sum_{i=1}^v d_0(s_i, t_{\pi_0(i)})$, it is enough to show that $d_1(\varrho^{(j-1)},\varrho^{(j)}) = \frac{1}{v} d_0(s_j, t_{\pi_0(j)})$. This follows from the more general fact that for two point measures which are equal except for the positions of one point in each measure, a minimizing assignment is given by pairing all of the coinciding points and pairing the two points whose locations differ.

The proof of this last statement is very easy. Let $\tilde\varrho_1 = \sum_{i=1}^v \delta_{\tilde s_i}$ and $\tilde\varrho_2 = \sum_{i=1}^v \delta_{\tilde t_i}$ with $\tilde s_i = \tilde t_i$ for $i \ne v$, that is, we have two point measures of the above form, renumbered such that the differing points come last and with points paired in the way which we claim to be optimal. Let $\pi \in \Sigma_v$, and denote that cycle of π which contains v by $(i_1 = v, i_2, \dots, i_l)$; additionally, set $i_{l+1} := v$. Since $\tilde s_i = \tilde t_i$ for $i \ne v$, we then have

$$d_0(\tilde s_v, \tilde t_v) \le \sum_{r=1}^{l} d_0(\tilde s_{i_r}, \tilde t_{i_{r+1}}),$$

and thus

$$\frac{1}{v}\sum_{i=1}^{v} d_0(\tilde{s}_i, \tilde{t}_i) = \frac{1}{v} d_0(\tilde{s}_v, \tilde{t}_v)$$
$$\leq \frac{1}{v}\sum_{i=1}^{v} d_0(\tilde{s}_i, \tilde{t}_{\pi(i)}).$$

Since π gives an arbitrary pairing, this yields that our claimed pairing is optimal. □

Corollary 3.3.C. *Let $f : \mathfrak{N}' \to \mathbb{R}$ be a function with*

(i) $|f(\varrho_1) - f(\varrho_2)| \leq 1$ for all $\varrho_1, \varrho_2 \in \mathfrak{N}'$ with $|\varrho_1| \neq |\varrho_2|$;

(ii) $|f(\varrho_1) - f(\varrho_2)| \leq \frac{1}{v} d_0(s,t)$ for all $\varrho_1, \varrho_2 \in \mathfrak{N}'$ of the form $\varrho_1 = \sum_{i=1}^{v-1} \delta_{s_i} + \delta_s$, $\varrho_2 = \sum_{i=1}^{v-1} \delta_{s_i} + \delta_t$, where $v \in \mathbb{N}$ and $s_1, \ldots, s_{v-1}, s, t \in E'$.

Then $f \in \mathfrak{F}_2$.

Proof. We have to establish that

$$|f(\varrho_1) - f(\varrho_2)| \leq d_1(\varrho_1, \varrho_2) \tag{3.12}$$

for all $\varrho_1, \varrho_2 \in \mathfrak{N}'$. For $\varrho_1 = \varrho_2 = 0$, Inequality (3.12) is trivially satisfied, and for $\varrho_1, \varrho_2 \in \mathfrak{N}'$ with $|\varrho_1| \neq |\varrho_2|$, it follows from (i). For $\varrho_1 = \sum_{i=1}^{v} \delta_{s_i}, \varrho_2 = \sum_{i=1}^{v} \delta_{t_i} \in \mathfrak{N}'$, we have in the notation of Lemma 3.3.B, using a minimizing pairing $\pi_0 \in \Sigma_v$,

$$|f(\varrho_1) - f(\varrho_2)| \leq \sum_{j=1}^{v} \left| f(\varrho^{(j-1)}) - f(\varrho^{(j)}) \right|$$
$$\leq \sum_{j=1}^{v} d_1(\varrho^{(j-1)}, \varrho^{(j)})$$
$$= d_1(\varrho_1, \varrho_2),$$

where the second inequality is due to (ii). Thus, we have established Inequality (3.12) for all $\varrho_1, \varrho_2 \in \mathfrak{N}'$. □

For greater clarity of the subsequent results and proofs in this chapter, we make two conventions. First, all function values of any $f : \mathfrak{N}' \to \mathbb{R}$ that are either not explicitly defined or that are defined in the context of a general formula which for the particular argument makes no sense, are to be understood as being zero. By this convention, the function f from Lemma 3.3.E, for example, is given by the stated formula for $|\varrho| \geq l$, and by $f(\varrho) = 0$ for $|\varrho| < l$.

The second convention is that we make a strict distinction between points of a point measure (denoted by s or t) and general locations in E' (denoted by x or y), to the effect that two "different" points s and t of a point measure can be at the same location $x \in E'$. This makes it more convenient to formulate arguments for multiple points in everyday language than if one has to distinguish points by their indices. From a formal point of view, we still write $s = t = x$.

The following two lemmas describe a rather general class of $\mathfrak{F}_2$-functions (see also Barbour, Holst, and Janson (1992), Section 10.2, for the first lemma). Essentially, these are statistics of point patterns that average over spatial quantities, especially ones given in terms of d_0-distances, for subpatterns of points.

Lemma 3.3.D. *Let $K : \mathbb{Z}_+ \times (E')^l \to [0,1]$ be a function that satisfies*

$$\big|K(u, x_1, \dots, x_l) - K(u, y_1, \dots, y_l)\big| \le \frac{1}{l} \sum_{i=1}^{l} d_0(x_i, y_i)$$

for all $u \in \mathbb{Z}_+$ and $x_1, \dots, x_l, y_1, \dots, y_l \in E'$, and define $f : \mathfrak{N}' \to [0,1]$ by

$$f(\varrho) := \frac{1}{|\varrho|^l} \int_{(E')^l} K(|\varrho|; x_1, \dots, x_l)\, \varrho(dx_1) \dots \varrho(dx_l) = \frac{1}{v^l} \sum_{i_1, \dots, i_l = 1}^{v} K(v; s_{i_1}, \dots, s_{i_l})$$

for $\varrho = \sum_{i=1}^{v} \delta_{s_i} \in \mathfrak{N}'$. Then $f \in \mathfrak{F}_2$.

For considering the second lemma, recall from Section 2.1 that we attach the symbol "$<$" to a summation sign in order to indicate that the summation indices are to be chosen in strictly increasing order. Hence, $\sum_{i_1, \dots, i_l = 1}^{v, <}$ below denotes summation over $i_1, \dots, i_l \in \{1, \dots, v\}$ with $i_1 < \dots < i_l$.

Lemma 3.3.E. *Let $K : \mathbb{Z}_+ \times (E')^l \to [0,1]$ be a symmetric function that satisfies*

$$\big|K(u, x_1, \dots, x_l) - K(u, y_1, \dots, y_l)\big| \le \frac{1}{l} \sum_{i=1}^{l} d_0(x_i, y_i) \tag{3.13}$$

for all $u \in \mathbb{Z}_+$ and $x_1, \dots, x_l, y_1, \dots, y_l \in E'$, and define $f : \mathfrak{N}' \to [0,1]$ by

$$f(\varrho) := \frac{1}{\binom{v}{l}} \sum_{i_1, \dots, i_l = 1}^{v, <} K(v; s_{i_1}, \dots, s_{i_l}) \tag{3.14}$$

for $\varrho = \sum_{i=1}^{v} \delta_{s_i} \in \mathfrak{N}'$. Then $f \in \mathfrak{F}_2$.

Proof. We give only the proof of Lemma 3.3.E, because we use only this result later on and the proof of Lemma 3.3.D is almost identical. Let $\varrho_1, \varrho_2 \in \mathfrak{N}'$. We show that $|f(\varrho_1) - f(\varrho_2)| \le d_1(\varrho_1, \varrho_2)$. If $|\varrho_1| \ne |\varrho_2|$, this follows from $0 \le f \le 1$, which in turn is a consequence of $0 \le K \le 1$. Suppose now that $\varrho_1 = \sum_{i=1}^{v} \delta_{s_i}$ and $\varrho_2 = \sum_{i=1}^{v} \delta_{t_i}$, where $v \ge l$ and where the points are numbered according to a minimizing pairing, so that we have $\frac{1}{v} \sum_{i=1}^{v} d_0(s_i, t_i) = d_1(\varrho_1, \varrho_2)$. Then

$$\begin{aligned}
\big|f(\varrho_1) - f(\varrho_2)\big| &\le \frac{1}{\binom{v}{l}} \sum_{i_1, \dots, i_l = 1}^{v, <} \big|K(v; s_{i_1}, \dots, s_{i_l}) - K(v; t_{i_1}, \dots, t_{i_l})\big| \\
&\le \frac{1}{l} \frac{1}{\binom{v}{l}} \sum_{i_1, \dots, i_l = 1}^{v, <} \sum_{r=1}^{l} d_0(s_{i_r}, t_{i_r}) \\
&= \frac{1}{l} \frac{1}{\binom{v}{l}} \binom{v-1}{l-1} \sum_{i=1}^{v} d_0(s_i, t_i) \\
&= d_1(\varrho_1, \varrho_2),
\end{aligned}$$

where the third line is obtained, because each $i \in \{1, \dots, v\}$ appears in exactly $\binom{v-1}{l-1}$ of the sets $\{i_1, \dots, i_l\} \subset \{1, \dots, v\}$ with l different elements. Altogether, we have $f \in \mathfrak{F}_2$. □

Some of the more concrete functions below are of the form given in Lemma 3.3.E. Others, like the average nearest neighbor distance, are not exactly of this form, but "are in the same spirit".

We now present the list of $\mathfrak{F}_2$-functions. An overview is given in Table 3.3.1. Note that we list d_1-Lipschitz continuous functions in general, so that the actual $\mathfrak{F}_2$-functions are obtained by dividing the functions in the table by their Lipschitz constants.

Name of function	Lipschitz constant
1) Indicator of total point count	1
2) Average location-to-point distance	1
3) Average point-to-point distance	2
4) Average distance to centroid for l-groups in $\mathbb{R}^D$	$2(l-1)/l$
5) Kernel estimator for a probability density at $x \in \mathbb{R}^D$	c/h^{D+1}
6) Mean absolute deviation of point-to-point distances	4
7) Average nearest-neighbor distance in $\mathbb{R}^D$	$\tau_D + 1$
8) Average r-scan in $\mathbb{R}$	1
9) Average union set volume per particle in $\mathbb{R}^D$	$2D$

TABLE 3.3.1: SOME LIPSCHITZ CONTINUOUS FUNCTIONS OF POINT MEASURES. For the meaning of the quantities c, h, and τ_D, as well as for the exact definition of the functions and the proof that they are Lipschitz continuous, see below.

In what follows, we discuss each example in detail. We always require that $\varrho = \sum_{i=1}^{v} \delta_{s_i} \in \mathfrak{N}'$.

1) Indicator of total point count

Let $A \subset \mathbb{Z}_+$ and

$$f_A(\varrho) := \mathrm{I}[|\varrho| \in A].$$

Then $f_A \in \mathfrak{F}_2$, because, for all $\varrho_1, \varrho_2 \in \mathfrak{N}'$, $\big|f_A(\varrho_1) - f_A(\varrho_2)\big| = 0$ if $|\varrho_1| = |\varrho_2|$, and $\big|f_A(\varrho_1) - f_A(\varrho_2)\big| \le 1 = d_1(\varrho_1, \varrho_2)$ if $|\varrho_1| \neq |\varrho_2|$. Note that the fact that $f_A \in \mathfrak{F}_2$ for every $A \subset \mathbb{Z}_+$ implies that

$$d_{TV}\big(\mathscr{L}(|\xi|), \mathscr{L}(|\eta|)\big) = \sup_{A \subset \mathbb{Z}_+} \big|\mathbb{E}(f_A(\xi)) - \mathbb{E}(f_A(\eta))\big| \le d_2\big(\mathscr{L}(\xi), \mathscr{L}(\eta)\big) \tag{3.15}$$

for any point processes ξ and η on E'.

2) Average location-to-point distance

Let $x_0 \in E'$ and

$$f_{x_0}(\varrho) := \frac{1}{v} \sum_{i=1}^{v} d_0(x_0, s_i).$$

Then $f_{x_0} \in \mathfrak{F}_2$, because f_{x_0} is of the form given in Equation (3.14) of Lemma 3.3.E with $l = 1$ and $K(u; x) := d_0(x_0, x)$, where the condition in (3.13) is satisfied by the inverse triangle inequality.

3) Average point-to-point distance

Let

$$f(\varrho) := \frac{1}{\binom{v}{2}} \sum_{i,j=1}^{v,<} d_0(s_i, s_j).$$

Then f is d_1-Lipschitz continuous with constant 2, or equivalently, $\frac{1}{2}f \in \mathfrak{F}_2$, because $\frac{1}{2}f$ is of the form given in Lemma 3.3.E with $l = 2$ and $K(u; x_1, x_2) := \frac{1}{2}d_0(x_1, x_2)$. Condition (3.13) for K follows by $|d_0(x_1, x_2) - d_0(y_1, y_2)| \le d_0(x_1, y_1) + d_0(x_2, y_2)$, which in turn is a consequence of the triangle inequality for d_0.

4) Average distance to centroid for l-groups in $\mathbb{R}^D$:

Let E' be a subset of $\mathbb{R}^D$, and let d_0 be any metric on E' that is induced by a norm $\|\cdot\|$ on $\mathbb{R}^D$ and satisfies $\operatorname{diam}(E') \le 1$. Fix $l \ge 2$. For locations $x_1, \dots, x_l \in E'$, denote by $g(x_1, \dots, x_l) := \frac{1}{l}\sum_{i=1}^{l} x_i$ their geometric centroid (i.e. their center of gravity). Set then

$$f_l(\varrho) := \frac{1}{\binom{v}{l}} \sum_{i_1,\dots,i_l=1}^{v,<} \overline{G}(s_{i_1}, \dots, s_{i_l}),$$

where $\overline{G}(x_1, \dots, x_l) := \frac{1}{l}\sum_{i=1}^{l} d_0\big(x_i, g(x_1, \dots, x_l)\big)$ denotes the average distance of the locations $x_1, \dots, x_l$ from their centroid. Then f is d_1-Lipschitz continuous with constant $2(l-1)/l$.

Again we show that up to the factor $2(l-1)/l$ the function f is of the form from Lemma 3.3.E, where this time $K(u, \cdot) := \frac{l}{2(l-1)}\overline{G}(\cdot)$ for any $u \in \mathbb{Z}_+$. We have $0 \le \overline{G} \le 1$, and hence $0 \le \frac{l}{2(l-1)}\overline{G} \le 1$, as required. So it remains to be shown that $\frac{l}{2(l-1)}\overline{G}$ satisfies Inequality (3.13).

For any $x_1, \dots, x_{l-1}, x, y, z \in E'$, we have

$$\begin{aligned}\Big|d_0\big(z, g(x_1, \dots, x_{l-1}, x)\big) - d_0\big(z, g(x_1, \dots, x_{l-1}, y)\big)\Big| \\ \le d_0\big(g(x_1, \dots, x_{l-1}, x), g(x_1, \dots, x_{l-1}, y)\big) = \left\|\frac{1}{l}(x-y)\right\| = \frac{1}{l}d_0(x, y)\end{aligned}$$

and

$$\begin{aligned}\Big|d_0\big(x, g(x_1, \dots, x_{l-1}, x)\big) - d_0\big(y, g(x_1, \dots, x_{l-1}, y)\big)\Big| \\ \le \left\|\left(x - \frac{1}{l}\left(\sum_{i=1}^{l-1} x_i + x\right)\right) - \left(y - \frac{1}{l}\left(\sum_{i=1}^{l-1} x_i + y\right)\right)\right\| \\ = \left\|\frac{l-1}{l}(x-y)\right\| = \frac{l-1}{l}d_0(x, y).\end{aligned}$$

Therefore,

$$\begin{aligned}
&\Big|\overline{G}(x_1,\dots,x_{l-1},x)-\overline{G}(x_1,\dots,x_{l-1},y)\Big| \\
&\qquad\le \frac{1}{l}\Bigg(\sum_{i=1}^{l-1}\Big|d_0\big(x_i,g(x_1,\dots,x_{l-1},x)\big)-d_0\big(x_i,g(x_1,\dots,x_{l-1},y)\big)\Big| \\
&\qquad\qquad\qquad +\Big|d_0\big(x,g(x_1,\dots,x_{l-1},x)\big)-d_0\big(y,g(x_1,\dots,x_{l-1},y)\big)\Big|\Bigg) \\
&\qquad\le \frac{1}{l}\Bigg(\sum_{i=1}^{l-1}\frac{1}{l}d_0(x,y)+\frac{l-1}{l}d_0(x,y)\Bigg) \\
&\qquad\le \frac{2(l-1)}{l^2}d_0(x,y).
\end{aligned}$$

Since $\overline{G}$ is symmetric, this yields

$$\begin{aligned}
&\big|\overline{G}(x_1,\dots,x_l)-\overline{G}(y_1,\dots,y_l)\big| \\
&\qquad\le \sum_{i=1}^{l}\Big|\overline{G}(x_1,\dots,x_{i-1},x_i,y_{i+1},\dots,y_l)-\overline{G}(x_1,\dots,x_{i-1},y_i,y_{i+1},\dots,y_l)\Big| \\
&\qquad\le \frac{2(l-1)}{l}\frac{1}{l}\sum_{i=1}^{l}d_0(x_i,y_i) \qquad (3.16)
\end{aligned}$$

for all $x_1,\dots,x_l,y_1,\dots,y_l\in E'$, whence it follows that $\frac{l}{2(l-1)}\overline{G}$ satisfies Inequality (3.13).

5) Kernel estimator for a probability density at $x\in\mathbb{R}^D$

Suppose that $E'\subset\mathbb{R}^D$ and that d_0 is the Euclidean metric truncated at 1. Let $\widetilde{K}:\mathbb{R}^D\to\mathbb{R}_+$ be a d_0-Lipschitz continuous function with constant c that satisfies $\int\widetilde{K}(s)\,ds=1$. Set

$$f_{x_0,h}(\varrho):=\frac{1}{vh^D}\sum_{i=1}^{v}\widetilde{K}\Big(\frac{s_i-x_0}{h}\Big)$$

for any $h\in(0,1]$ and any $x_0\in E'$. Then f is d_1-Lipschitz continuous with constant c/h^{D+1}, which follows from Lemma 3.3.E, where $K(u,x):=\frac{h}{c}\widetilde{K}\big((x-x_0)/h\big)$ for any $u\in\mathbb{Z}_+$ and any $x\in E'$.

The function $f_{x_0,h}$ has the following application. Suppose ϱ consists of data points obtained by independent sampling from a probability distribution with unknown density g. We then can interprete $f_{x_0,h}$ as the kernel estimator of $g(x_0)$ for the kernel $\widetilde{K}$ at window width h.

A very similar application is considered in much more detail in Subsection 5.3.2.

6) Mean absolute deviation of point-to-point distances

This is the first of several functions that do not exactly fit into the paradigm of Lemmas 3.3.D and 3.3.E, but have a similar concept. Suppose that E' is again a general

compact space, and let $g(\varrho)$ be the average point-to-point distance of ϱ as introduced in Example 3. Set

$$f(\varrho) := \frac{1}{\binom{v}{2}} \sum_{i,j=1}^{v,<} \big| d_0(s_i, s_j) - g(\varrho) \big|.$$

Then f is d_1-Lipschitz continuous with constant 4.

We use Corollary 3.3.C to show this statement. The requirement (i) from that corollary is satisfied for $\frac{1}{4}f$, because $0 \le f \le 1$. We show (ii) for $v \ge 2$. Let $\varrho_1 = \sum_{i=1}^{v} \delta_{s_i}$ and $\varrho_2 = \sum_{i=1}^{v} \delta_{t_i}$ with $s_i = t_i$ for $1 \le i \le v-1$, and set $s := s_v$ and $t := t_v$. Then

$$\begin{aligned}
\big| f(\varrho_1) - f(\varrho_2) \big| &\le \frac{1}{\binom{v}{2}} \sum_{i,j=1}^{v,<} \Big| \big| d_0(s_i, s_j) - g(\varrho_1) \big| - \big| d_0(t_i, t_j) - g(\varrho_2) \big| \Big| \\
&\le \frac{1}{\binom{v}{2}} \sum_{i,j=1}^{v,<} \Big(\big| d_0(s_i, s_j) - d_0(t_i, t_j) \big| + \big| g(\varrho_1) - g(\varrho_2) \big| \Big) \\
&= \frac{1}{\binom{v}{2}} \sum_{i=1}^{v-1} \big| d_0(s_i, s) - d_0(s_i, t) \big| \; + \; \big| g(\varrho_1) - g(\varrho_2) \big| \\
&\le \frac{2}{v} d_0(s,t) + \frac{2}{v} d_0(s,t) \\
&= \frac{4}{v} d_0(s,t),
\end{aligned}$$

where the first summand in the third line is obtained by $d_0(s_i, s_j) = d_0(t_i, t_j)$ for all $i, j \in \{1, \dots, v-1\}$, and the second summand in the fourth line by $\frac{1}{2}g \in \mathfrak{F}_2$. Thus, $\frac{1}{4}f$ satisfies also Property (ii) of Corollary 3.3.C.

7) Average nearest-neighbor distance in $\mathbb{R}^D$

Suppose that $E' \subset \mathbb{R}^D$ and d_0 is the Euclidean metric truncated at 1. Define the *average nearest-neighbor distance* of ϱ to be

$$f(\varrho) := \frac{1}{v} \sum_{i=1}^{v} \min_{\substack{j \in \{1,\dots,v\} \\ j \ne i}} d_0(s_i, s_j).$$

The statistic $f(\varrho)$ contains valuable information about the amount of clustering in the point pattern ϱ, at least when combined with information about the total number of points in the pattern (and the size and shape of the state space E'). An application of this statistic is given in Subsection 5.3.3.

We claim that f is d_1-Lipschitz continuous with constant $\tau_D + 1$, where τ_D is the maximal kissing number in D dimensions; that is, the highest number of unit balls (in the untruncated Euclidean metric) in $\mathbb{R}^D$ that can touch a given unit ball without any intersections of the interiors of the balls involved. The number τ_D is known explicitly only for dimensions D up to 4, for $D = 8$, and $D = 24$. The case $D = 4$ was only recently solved by Musin (2003). Table 3.3.2 presents an overview of known upper and lower bounds of τ_D for $D \le 10$.

D	τ_D	D	τ_D
1	2	6	72–82
2	6	7	126–140
3	12	8	240
4	24	9	306–380
5	40–46	10	500–595

TABLE 3.3.2: RANGES OF POSSIBLE VALUES FOR MAXIMAL KISSING NUMBERS IN LOW DIMENSIONS.

A rather obvious, but very rough quantitative upper bound is given by $\tau_D \leq 3^D - 1$, derived by arranging the centers of unit balls in $\{-2, 0, 2\}^D$. A much subtler qualitative upper bound is given in Kabatiansky and Levenshtein (1978) as $\tau_D \leq 2^{0.401D(1+o(1))}$. More details about maximal kissing numbers, and also a more complete table with upper and lower bounds for τ_D in dimensions up to 24 can be found in Conway and Sloane (1999), Section 1.2.

In order to show the required Lipschitz continuity of f, we again make use of Corollary 3.3.C. Property (i) from that corollary is satisfied for $(\tau_D+1)^{-1}f$, because $0 \leq f \leq 1$. In order to show Property (ii) for $v \geq 2$, let $\varrho_1 = \sum_{i=1}^{v} \delta_{s_i}$, $\varrho_2 = \sum_{i=1}^{v} \delta_{t_i}$ with $s_i = t_i$ for $1 \leq i \leq v-1$, and set $s := s_v$ and $t := t_v$. Then

$$|f(\varrho_1) - f(\varrho_2)| \leq \frac{1}{v}\left|\sum_{i=1}^{v-1}\left(\min_{\substack{1\leq j\leq v\\ j\neq i}} d_0(s_i, s_j) - \min_{\substack{1\leq j\leq v\\ j\neq i}} d_0(t_i, t_j)\right)\right| + \left|\min_{1\leq j\leq v-1} d_0(s, s_j) - \min_{1\leq j\leq v-1} d_0(t, t_j)\right|. \tag{3.17}$$

The differences on the right hand side of Equation (3.17) are zero unless they involve at least one of the points s and t, in which case they are at most $d_0(s,t)$. In detail, we have the following estimates. First,

$$\left|\min_{1\leq j\leq v-1} d_0(s, s_j) - \min_{1\leq j\leq v-1} d_0(t, t_j)\right| \leq d_0(s,t). \tag{3.18}$$

In order to see this, suppose that the two minima are realized in s^* and t^*, respectively, and that without loss of generality $d_0(s, s^*) > d_0(t, t^*)$ (otherwise switch s and t). Then

$$\left|d_0(s, s^*) - d_0(t, t^*)\right| = d_0(s, s^*) - d_0(t, t^*) \leq d_0(s, t^*) - d_0(t, t^*) \leq d_0(s, t),$$

where the inequality in the middle holds, because s^* and t^* are both chosen among the points $s_1, \ldots, s_{v-1}$, but s^* is closest to s among these points. Next, we have

$$\min_{\substack{1\leq j\leq v\\ j\neq i}} d_0(s_i, s_j) - \min_{\substack{1\leq j\leq v\\ j\neq i}} d_0(t_i, t_j) = \min_{\substack{1\leq j\leq v-1\\ j\neq i}} d_0(s_i, s_j) - \min_{\substack{1\leq j\leq v-1\\ j\neq i}} d_0(t_i, t_j) = 0 \tag{3.19}$$

for any $i \in \{1, \ldots, v-1\}$ for which neither s nor t is nearer to s_i than are all the other $s_j = t_j$ (where $1 \leq j \leq v-1$, $j \neq i$). Finally,

$$\left|\min_{\substack{1\leq j\leq v\\ j\neq i}} d_0(s_i, s_j) - \min_{\substack{1\leq j\leq v\\ j\neq i}} d_0(t_i, t_j)\right| \leq d_0(s,t) \tag{3.20}$$

for any $i \in \{1, \ldots, v-1\}$ for which either s or t is nearer to s_i than the other s_j are. To see this, suppose that that the minima are realized in s_i^* and t_i^*, respectively. If $s_i^* = s$ *and* $t_i^* = t$, then Inequality (3.20) is immediate by the inverse triangle inequality. Suppose therefore without loss of generality that $s_i^* \neq s$, but $t_i^* = t$. Then

$$\begin{aligned}\left|d_0(s_i, s_i^*) - d_0(t_i, t_i^*)\right| &= \left|d_0(s_i, s_i^*) - d_0(s_i, t)\right| \\ &= d_0(s_i, s_i^*) - d_0(s_i, t) \leq d_0(s_i, s) - d_0(s_i, t) \leq d_0(s, t),\end{aligned}$$

where the second relation holds, because s_i^* is one of the points t_j, $j \neq i$, from which the nearest neighbor t of $t_i = s_i$ is selected; and the third relation holds, because s is one of the points from which the nearest neighbor s_i^* of s_i is selected.

One can prove (see Zeger and Gersho (1994), Theorem 1) that any point $\tilde{s}$ of a point measure in $\mathbb{R}^D$ can be a nearest neighbor of at most τ_D other points, not counting points that share the same location with $\tilde{s}$. Hence, at most τ_D of the differences $\min_{j\neq i} d_0(s_i, s_j) - \min_{j\neq i} d_0(t_i, t_j)$, $1 \leq i \leq v-1$, are positive (namely the ones for which t is nearer to $t_i = s_i$ than the other $t_j = s_j$ are), and at most τ_D are negative (namely the ones for which s is nearer to $s_i = t_i$ than the other $s_j = t_j$ are). Note that this statement is also true if there are other points at the same location as s or t.

In total we have that, among the differences on the right hand side of Inequality (3.17), there are at most $2\tau_D + 1$ that are non-zero, of which at most $\tau_D + 1$ share the same sign. This together with Inequalities (3.18) and (3.20) yields the estimate

$$|f(\varrho_1) - f(\varrho_2)| \leq (\tau_D + 1)\frac{1}{v} d_0(s, t).$$

Thus, $(\tau_D + 1)^{-1} f$ satisfies also Property (ii) of Corollary 3.3.C, whence follows the Lipschitz continuity claimed.

8) Average r-scan in $\mathbb{R}$

Suppose that $E' \subset \mathbb{R}$ and d_0 is the Euclidean metric truncated at 1. Define

$$f_{r,h}(\varrho) := \frac{1}{v} \int_{E'} \mathrm{I}[\varrho(R_x) \geq r]\, dx$$

for $r \in \mathbb{N}$, $h \in (0, 1]$, and $R_x := [x, x + h)$. We call $f_{r,h}$ the *average r-scan function* for a scan interval of length h. Like the nearest-neighbor distance, the average r-scans of a point pattern contain important information about its structure. Typically, E' is an interval, and some kind of edge correction is applied (e.g. by integrating only over $[0, 1-h]$ instead of $E' = [0, 1]$). A relatively high value of $f_{r,h}(\varrho)$ for fixed r and h (it can easily be seen that $0 \leq f \leq h/r$) is then an indication that a rather large portion of the point pattern essentially consists of well-separated clusters of size r or a bit bigger, each with a span of an order of magnitude $\leq h$.

We claim that $f \in \mathfrak{F}_2$. This is shown by using once more Corollary 3.3.C. Property (i) of this corollary holds, because $0 \leq f \leq h/r \leq 1$. In order to show Property (ii) for $v \geq 1$, let $\varrho_1 = \sum_{i=1}^{v-1} \delta_{s_i} + \delta_s$ and $\varrho_2 = \sum_{i=1}^{v-1} \delta_{s_i} + \delta_t$, where without loss of generality $s < t$ (for

$s = t$ nothing has to be shown, and for $s > t$ just switch ϱ_1 and ϱ_2). Then

$$\begin{aligned}\big|f(\varrho_1) - f(\varrho_2)\big| &= \frac{1}{v}\bigg|\int_{E'} \big(\mathrm{I}[\varrho_1(R_x) \geq r] - \mathrm{I}[\varrho_2(R_x) \geq r]\big)\, dx\bigg| \\ &= \frac{1}{v}\bigg|\int_{(s-h,\, s\wedge(t-h)]\cap E'} \big(\mathrm{I}[\varrho_1(R_x) \geq r] - \mathrm{I}[\varrho_2(R_x) \geq r]\big)\, dx \\ &\qquad - \int_{((t-h)\vee s,\, t]\cap E'} \big(\mathrm{I}[\varrho_2(R_x) \geq r] - \mathrm{I}[\varrho_1(R_x) \geq r]\big)\, dx\bigg|. \qquad (3.21)\end{aligned}$$

The second equality holds by the following argument. The difference $\mathrm{I}[\varrho_1(R_x) \geq r] - \mathrm{I}[\varrho_2(R_x) \geq r]$ is $+1$ at most on the interval $(s-h, s\wedge(t-h)]$, because that is where the interval R_x sees the additional point s of ϱ_1 without seeing the corresponding point t of ϱ_2, and it is -1 at most on the interval $((t-h)\vee s, t]$, because that is where the interval R_x sees the additional point t of ϱ_2 without seeing the corresponding point s of ϱ_1. Otherwise R_x sees the same number of points of ϱ_1 as it does of ϱ_2, so that the difference is zero.

Note that both integrands on the right hand side of Equation (3.21) take only 0 and 1 as their values. Estimating the integrand of the larger integral by 1 and the integrand of the smaller integral by 0, yields that

$$\big|f(\varrho_1) - f(\varrho_2)\big| \leq \frac{1}{v}\min(h, t-s) \leq \frac{1}{v} d_0(s,t),$$

since $h \leq 1$. Thus, Property (ii) is established, and we obtain by Corollary 3.3.C that $f \in \mathfrak{F}_2$.

9) Average union set volume per particle in $\mathbb{R}^D$

Let $E' := \mathcal{K}'$ be the space of all non-empty closed convex subsets of $[0,1]^D \subset \mathbb{R}^D$. We call such a subset a *convex body* in $[0,1]^D$. As usual, we equip $[0,1]^D$ with the Euclidean metric truncated at 1. Let furthermore d_0 be the Hausdorff metric d_H on E', which is given by

$$d_H(K,L) := \max\Big(\max_{x\in K}\min_{y\in L}\big(|x-y|\wedge 1\big), \max_{y\in L}\min_{x\in K}\big(|x-y|\wedge 1\big)\Big)$$

for any $K, L \in \mathcal{K}'$. It is easily seen that d_H is a metric that is bounded by 1 and can be equivalently defined by

$$d_H(K,L) = \min\big\{\varepsilon \geq 0;\, K \subset L^{[\varepsilon]}, L \subset K^{[\varepsilon]}\big\} \qquad (3.22)$$

for all $K, L \in \mathcal{K}'$, where $F^{[\varepsilon]}$ denotes the *closed ε-halo set* of F for any closed subset F of $[0,1]^D$, that is

$$F^{[\varepsilon]} := \big\{y \in [0,1]^D;\, |y-x|\wedge 1 \leq \varepsilon \text{ for some } x \in F\big\}.$$

By Schneider and Weil (2000), Satz 1.2.1, it follows that d_0 metrizes the restriction to $E' = \mathcal{K}'$ of the hit-or-miss topology introduced in Subsection 2.3.1, which is the natural topology we would like to use. Furthermore $E' = \mathcal{K}'$ is compact, because it is a closed

subset of the system $\mathcal{F}$ of closed subsets of $[0,1]^D$, which is compact by Satz 2.1.2 in Schneider and Weil (2000). The fact that $\mathcal{K}'$ is a closed subset of $\mathcal{F}$ is easily derived from the characterization of the convergence in the hit-or-miss topology given as Satz 1.1.2 of the same book.

Hence the space (E', d_0) is both natural in view of the theory of stochastic geometry and suitable for the definition of the d_1- and d_2-metrics.

We write here $\varrho = \sum_{i=1}^{v} \delta_{K_i}$ instead of $\varrho = \sum_{i=1}^{v} \delta_{s_i}$ for our general point measure on E', and set

$$f(\varrho) := \frac{1}{v} Leb^D\Big(\bigcup_{i=1}^{v} K_i\Big), \tag{3.23}$$

which gives the average contribution of each particle of ϱ to the volume of the union set. This statistic gives valuable information about the structure of the union set. Especially if the average volume of the particles is also known, it describes to what degree the particles overlap.

We claim that f is Lipschitz continuous with constant $2D$. This is shown once more by Corollary 3.3.C. Property (i) holds for $(2D)^{-1}f$, because obviously $0 \le f \le 1$. In order to show Property (ii) for $v \le 1$, let $\varrho_1 = \sum_{i=1}^{v-1} \delta_{K_i} + \delta_K$ and $\varrho_2 = \sum_{i=1}^{v-1} \delta_{K_i} + \delta_L$ with $K_1, \ldots, K_{v-1}, K, L \in \mathcal{K}'$. Assume that $f(\varrho_1) - f(\varrho_2) \ge 0$ (switch ϱ_1 and ϱ_2 otherwise), and that $\varepsilon := d_H(K,L) < 1$ (Property (ii) is trivially satisfied if $d_H(K,L) = 1$). Since $K \subset L^{[\varepsilon]}$ by Equation (3.22), we then have

$$\begin{aligned}
\big|f(\varrho_1) - f(\varrho_2)\big| &\le \frac{1}{v}\bigg(Leb^D\Big(\bigcup_{i=1}^{v-1} K_i \cup L^{[\varepsilon]}\Big) - Leb^D\Big(\bigcup_{i=1}^{v-1} K_i \cup L\Big)\bigg) \\
&= \frac{1}{v} Leb^D\bigg(\Big(\bigcup_{i=1}^{v-1} K_i \cup L^{[\varepsilon]}\Big) \setminus \Big(\bigcup_{i=1}^{v-1} K_i \cup L\Big)\bigg) \\
&\le \frac{1}{v} Leb^D\big(L^{[\varepsilon]} \setminus L\big) \\
&\le 2D\frac{1}{v}\varepsilon = 2D\frac{1}{v} d_0(K,L).
\end{aligned} \tag{3.24}$$

The last inequality follows by a simple argument from convex geometry. By Satz 2.2.2 in Schneider and Weil (1992), the surface area $\widetilde{V}_{n-1}$ of a convex body (which is twice the $(n-1)$-th intrinsic volume V_{n-1} used in Schneider and Weil's book) is increasing with respect to set inclusion. Hence

$$\widetilde{V}_{n-1}\big(L^{[\varepsilon]}\big) \le \widetilde{V}_{n-1}\big([0,1]^D\big) = 2D.$$

This yields that

$$Leb^D\big(L^{[\varepsilon]} \setminus L\big) \le \widetilde{V}_{n-1}\big(L^{[\varepsilon]}\big)\varepsilon \le 2D\varepsilon,$$

where the first inequality follows by the convexity of $L^{[\varepsilon]}$. This is intuitively clear and can be formally proved by the approximation of the convex body $L^{[\varepsilon]}$ by convex polytopes, using that the ε-eroded set $\{y \in L^{[\varepsilon]};\, \mathbb{B}(y,\varepsilon) \subset L^{[\varepsilon]}\}$ of $L^{[\varepsilon]}$ is equal to L, and that the surface area $\widetilde{V}_{n-1}$ and the volume Leb^D are continuous functionals, which follows once more by Satz 2.2.2 in Schneider and Weil (1992).

Thus, by Inequality (3.24), $(2D)^{-1}f$ satisfies also Property (ii) from Corollary 3.3.C, whence follows the Lipschitz continuity claimed.

Here, as in some of the previous examples, we might prefer to dispose of the factor $\frac{1}{v}$ in the definition of the function f. Obviously, this is not possible if we want to retain a Lipschitz continuous function. Nevertheless, the fact that f as given in (3.23) is Lipschitz continuous implies the following result about the volume fraction of the union set of a particle process whose particles are convex bodies.

Let ξ and η be stationary particle processes, concentrated on the space of non-empty compact convex subsets of $\mathbb{R}^D$, and denote by Ξ_ξ and Ξ_η their union sets, which by Proposition 2.3.C are RACS. Recall that we define the volume fraction p_Ξ of a stationary RACS by $p_\Xi = \mathbb{E}\big(Leb^D(\Xi \cap [0,1]^D)\big)$. Write ξ' and η' for the restrictions of the particle processes ξ and η to $[0,1]^D$, that is, the processes that are constructed by intersecting every particle with $[0,1]^D$ and leaving away any empty sets obtained in the process. We then have for any particle processes $\tilde{\xi} \stackrel{\mathscr{D}}{=} \xi'$ and $\tilde{\eta} \stackrel{\mathscr{D}}{=} \eta'$, and arbitrary $m \in \mathbb{N}$ that

$$\begin{aligned}\big|p_{\Xi_\xi} - p_{\Xi_\eta}\big| &= \Big|\mathbb{E}\big(|\tilde{\eta}|(f(\tilde{\xi}) - f(\tilde{\eta}))1_{\{|\tilde{\xi}|=|\tilde{\eta}|\}}\big) + \mathbb{E}\big((p_{\Xi_{\tilde{\xi}}} - p_{\Xi_{\tilde{\eta}}})1_{\{|\tilde{\xi}|\neq|\tilde{\eta}|\}}\big)\Big| \\ &\leq 2D\,m\,\mathbb{E}\big(d_1(\tilde{\xi},\tilde{\eta})1_{\{|\tilde{\xi}|=|\tilde{\eta}|\}}\big) + \mathbb{E}\big(|\tilde{\eta}|1_{\{|\tilde{\eta}|>m\}}\big) + \mathbb{E}\big(d_1(\tilde{\xi},\tilde{\eta})1_{\{|\tilde{\xi}|\neq|\tilde{\eta}|\}}\big),\end{aligned}$$

and hence, taking the minimum over all such $\tilde{\xi}$ and $\tilde{\eta}$,

$$\big|p_{\Xi_\xi} - p_{\Xi_\eta}\big| \leq 2D\,m\,d_2\big(\mathscr{L}(\xi'),\mathscr{L}(\eta')\big) + \mathbb{E}\big(|\eta'|1_{\{|\eta'|>m\}}\big) \tag{3.25}$$

for any $m \in \mathbb{N}$. Note that $|\eta'|$ is equal to $\eta(\mathcal{F}_{[0,1]^D})$, the number of particles of η that have a non-empty intersection with $[0,1]^D$. If η is a Poisson process with intensity measure λ, the second summand of Inequality (3.25) can therefore be estimated as in Inequality (3.10) by $(9\pi m)^{-1/2}e^{-m}$ for $m \geq e^2\lambda(\mathcal{F}_{[0,1]^D})$.

3.3.2 Other possibilities to make use of d_2 upper bounds

So far, our focus was on finding functions $f \in \mathfrak{F}_2$, for which we then can bound terms of the form $d_W\big(\mathscr{L}(f(\xi)),\mathscr{L}(f(\eta))\big)$ by $d_2(\mathscr{L}(\xi),\mathscr{L}(\eta))$ via Lemma 3.3.A. However, this is just a special case of bounding a term $\Phi(\mathscr{L}(\xi),\mathscr{L}(\eta))$, where $\Phi : \mathfrak{P}(\mathfrak{N}')^2 \to \mathbb{R}_+$, by a term $\phi\big(d_2(\mathscr{L}(\xi),\mathscr{L}(\eta))\big)$, where $\phi : \mathbb{R}_+ \to \mathbb{R}_+$, which might be of interest in situations that cannot be realized via $\mathfrak{F}_2$-functions. We present here the Wasserstein distance between the spherical contact distributions of two simple stationary point processes as one possible example for such a term $\Phi(\mathscr{L}(\xi),\mathscr{L}(\eta))$, and consider furthermore the Wasserstein distance between certain statistics that depend on these distributions. There are of course many more such examples, which typically can be dealt with in a somewhat similar way, but often require some extra considerations for the concrete situation.

In Subsection 2.2.9, the functions F, G, J, K, and L were introduced, which carry important information about the distribution of a given point process. Obviously, they are not functions on $\mathfrak{N}'$, but rather functions on $\mathfrak{P}(\mathfrak{N}')$ directly, so that they are not suited for the considerations via $\mathfrak{F}_2$-functions. We confine ourselves here to the empty space function F. In a first step we examine $d_W(F,F')$ for the empty space functions F and F' of two simple stationary point processes ξ and η on $\mathbb{R}^D$, where by a slight abuse

of notation, we write the distribution functions F and F' as arguments of d_W instead of the distributions which they characterize.

Unfortunately, it does not seem possible to obtain a direct upper bound for $d_W(F, F')$ in terms of the d_2-distance between the distributions of ξ and η; but we easily obtain their d_2^*-distance as an upper bound, which in turn can be bounded by an expression involving their d_2-distance via Propositions 3.2.B or 3.2.C.

Note that the F-function that we use here is not exactly the F-function from Subsection 2.2.9, but an approximation that has a jump at 1 to the function value 1. In other words, it is the F-function with respect to the d_0-distance $|\cdot| \wedge 1$ instead of the untrimmed Euclidean distance $|\cdot|$ on $\mathbb{R}^D$. Since the bound 1 on d_0 was chosen arbitrarily and since up to the additional factor $\gamma > 0$ the same metrics d_1 and d_2 are obtained when trimming the Euclidean metric at γ, we can get arbitrarily close approximations to the F-function from Subsection 2.2.9.

In order to simplify the notation, we denote in what follows by $\mathbb{B}(x, 1)$ the closed unit ball at $x \in \mathbb{R}^D$ in the Euclidean metric (note that $\mathbb{B}_0(x, 1) = \mathbb{R}^D$ for the closed unit ball in the d_0-metric).

Proposition 3.3.F. *Let ξ and η be simple stationary point processes on $\mathbb{R}^D$ with empty space functions F and F', respectively. We then have*

$$d_W(F, F') \le d_2^*\big(\mathscr{L}(\xi|_{\mathbb{B}(0,1)}), \mathscr{L}(\eta|_{\mathbb{B}(0,1)})\big).$$

Proof. Let $\tilde{\xi}$ and $\tilde{\eta}$ be arbitrary point processes on $\mathbb{B}(0, 1)$ with $\tilde{\xi} \stackrel{\mathscr{D}}{=} \xi|_{\mathbb{B}(0,1)}$ and $\tilde{\eta} \stackrel{\mathscr{D}}{=} \eta|_{\mathbb{B}(0,1)}$. Denote by $S_1, S_2, \ldots$ and $S_1', S_2', \ldots$ the points of $\tilde{\xi}$ and $\tilde{\eta}$, respectively, in the order of their distances to the origin (with an arbitrary secondary criterion if, for either of the point processes, the probability that two of these distances are the same is positive). Denote by Π the (random) permutation that gives the d_P-minimizing pairing between $S_1, \ldots, S_{|\tilde{\xi}|}$ and $S_1', \ldots, S_{|\tilde{\eta}|}'$ on $\{|\tilde{\xi}| = |\tilde{\eta}|\}$, so that $\max_{1 \le i \le |\tilde{\xi}|} d_0(S_i, S'_{\Pi(i)}) 1_{\{|\tilde{\xi}|=|\tilde{\eta}|\}} = d_P(\tilde{\xi}, \tilde{\eta}) 1_{\{|\tilde{\xi}|=|\tilde{\eta}|\}}$. Then

$$\begin{aligned} d_W(F, F') &\le \mathbb{E} \int_0^1 \big| 1_{\{\tilde{\xi}(\mathbb{B}(0,r)) \ge 1\}} - 1_{\{\tilde{\eta}(\mathbb{B}(0,r)) \ge 1\}} \big| \, dr \\ &\le \mathbb{E}\big(\big| |S_1| - |S_1'| \big| 1_{\{|\tilde{\xi}|=|\tilde{\eta}| \ge 1\}} \big) + \mathbb{E}\big(1_{\{|\tilde{\xi}| \ne |\tilde{\eta}|\}} \big) \\ &\le \mathbb{E}\big(\big| S_1 - S'_{\Pi(1)} \big| 1_{\{|\tilde{\xi}|=|\tilde{\eta}| \ge 1\}} \big) + \mathbb{E}\big(d_P(\tilde{\xi}, \tilde{\eta}) 1_{\{|\tilde{\xi}| \ne |\tilde{\eta}|\}} \big) \\ &\le \mathbb{E} d_P(\tilde{\xi}, \tilde{\eta}), \end{aligned} \tag{3.26}$$

where we use Equation (3.11) for the first inequality, and the second inequality is seen by noting that

$$\big| 1_{\{\tilde{\xi}(\mathbb{B}(0,r)) \ge 1\}} - 1_{\{\tilde{\eta}(\mathbb{B}(0,r)) \ge 1\}} \big| = 1_{\{r \in [|S_1| \wedge |S_1'|, \, |S_1| \vee |S_1'|)\}}.$$

Taking the minimum over all $\tilde{\xi} \stackrel{\mathscr{D}}{=} \xi$ and $\tilde{\eta} \stackrel{\mathscr{D}}{=} \eta$ in Inequality (3.26) yields the result claimed. □

Remark 3.3.G. Note that we also obtain by the same proof that

$$d_W(F, F') \le d_2^*\big(\mathscr{L}(\xi|_C), \mathscr{L}(\eta|_C)\big)$$

for every compact set $C \subset \mathbb{R}^D$ which contains $\mathbb{B}(0, 1)$.

In what follows, we present an application for the bound on $d_W(F, F')$, which needs some other ingredients as well. Suppose that one wants to test if an observed point pattern in a compact window $W \subset \mathbb{R}^D$ is a realization of a certain simple stationary point process ξ. An often-used technique is to estimate one or several of the distance-based functions F, G, J, K, and L for the observed point pattern, and then perform a goodness-of-fit test based on some measure of "function discrepancy" between the estimated and the corresponding theoretical function.

We explain this method in somewhat more detail for the F-function. An unbiased non-parametric estimate for the F-function based on a point pattern ϱ is given by

$$\widehat{F}(r) := \widehat{F}_\varrho(r) := \frac{1}{|W^{[-r]}|} \int_{W^{[-r]}} \mathrm{I}[\varrho(\mathbb{B}(x,r)) \geq 1]\, dx \quad \text{for } r \in [0,1) \tag{3.27}$$

where $W^{[-r]} := \{y \in W;\, \mathbb{B}(y,r) \subset W\}$ denotes the *r-eroded set* of W. We assume here that $W^{[-1]} \neq \emptyset$, in order to have $|W^{[-r]}| > 0$ for $r \in [0,1)$. For the sake of simplicity, we require the origin to lie in $W^{[-1]}$, so that $\mathbb{B}(0,1) \subset W$. The use of $W^{[-r]}$ instead of W in the definition of $\widehat{F}(r)$ constitutes the most simple possibility of a boundary correction, called *minus-sampling*, in order to make $\widehat{F}(r)$ unbiased. Note that, because of the minus-sampling, $\widehat{F}$ is in general not a distribution function, because it might be that it is not increasing over all of its domain. This, however, hardly ever poses any practical problems.

The term on the right hand side of Equation (3.27) is calculated via the approximation

$$\widehat{F}_\varrho(r) \approx \frac{1}{\#(W^{[-r]} \cap G)} \sum_{x \in W^{[-r]} \cap G} \mathrm{I}[\varrho(\mathbb{B}(x,r)) \geq 1], \tag{3.28}$$

where G is a finite regular grid in $\mathbb{R}^D$; see Møller and Waagepetersen (2004), Subsection 4.3.6 for details. In what follows, we ignore the difference between the expressions in (3.27) and (3.28), which can easily be controlled by the mesh size of the chosen grid.

If we interprete $\widehat{F}_\varrho(r)$ as a function in ϱ, it is very closely related to a multidimensional version of the average 1-scan from the previous subsection. By this analogy, the following result can be easily shown.

Proposition 3.3.H. *Let $E' := W$, $r \in (0,1)$, and let $\hat{f}_r : \mathfrak{N}' \to \mathbb{R}_+$ be defined by*

$$\hat{f}_r(\varrho) := \frac{|W^{[-r]}|}{|\varrho|} \widehat{F}_\varrho(r) \quad \textit{for } \varrho \in \mathfrak{N}' \setminus \{0\}.$$

Then $\hat{f}_r$ is d_1-Lipschitz continuous with constant $c := (\alpha_{D-1} r^{D-1}) \vee (\alpha_D r^D) \leq 2\alpha_{D-1} r^{D-1}$.

Proof. We proceed in a very similar way as in the proof for the average r-scan function; more detailed explanations can be found there. Property (i) of Corollary 3.3.C holds for $\frac{1}{c}\hat{f}_r$, because $0 \leq \hat{f}_r \leq \alpha_D r^D$. In order to show Property (ii), let $\varrho_1 = \sum_{i=1}^{v-1} \delta_{s_i} + \delta_s$

and $\varrho_2 = \sum_{i=1}^{v-1} \delta_{s_i} + \delta_t$ for $v \in \mathbb{N}$ and $s_1, \ldots, s_{v-1}, s, t \in E'$. Then

$$\begin{aligned}
\big|\hat{f}_r(\varrho_1) - \hat{f}_r(\varrho_2)\big| &= \frac{1}{v}\bigg|\int_{W^{[-r]}} \Big(\mathrm{I}\big[\varrho_1(\mathbb{B}(x,r)) \geq 1\big] - \mathrm{I}\big[\varrho_2(\mathbb{B}(x,r)) \geq 1\big]\Big)\, dx\bigg| \\
&= \frac{1}{v}\bigg|\int_{(\mathbb{B}(s,r)\setminus\mathbb{B}(t,r))\cap W^{[-r]}} \Big(\mathrm{I}\big[\varrho_1(\mathbb{B}(x,r)) \geq 1\big] - \mathrm{I}\big[\varrho_2(\mathbb{B}(x,r)) \geq 1\big]\Big)\, dx \\
&\qquad - \int_{(\mathbb{B}(t,r)\setminus\mathbb{B}(s,r))\cap W^{[-r]}} \Big(\mathrm{I}\big[\varrho_2(\mathbb{B}(x,r)) \geq 1\big] - \mathrm{I}\big[\varrho_1(\mathbb{B}(x,r)) \geq 1\big]\Big)\, dx\bigg| \\
&\leq \frac{1}{v} Leb^D\big(\mathbb{B}(t,r) \setminus \mathbb{B}(s,r)\big) \\
&\leq \frac{1}{v} Leb^D\Big(\big(\mathbb{B}(s,r) + [\![0, t-s]\!]\big) \setminus \mathbb{B}(s,r)\Big) \\
&= \frac{1}{v}\alpha_{D-1} r^{D-1} d_0(s,t) \qquad (3.29)
\end{aligned}$$

for $d_0(s,t) < 1$, where $[\![0, t-s]\!] := \{\delta(t-s);\, \delta \in [0,1]\}$. For $d_0(s,t) = 1$, we have $|\hat{f}_r(\varrho_1) - \hat{f}_r(\varrho_2)| \leq \alpha_D r^D = \alpha_D r^D d_0(s,t)$. Corollary 3.3.C now implies the required statement. □

One of the theoretically more sound ways to proceed for testing if an observed point pattern ϱ is a realization of the point process ξ with empty space function F_0 is to measure the overall discrepancy between the estimated function $\widehat{F}$ and the true function F_0, for example by

$$T_{F_0}(\varrho) := \int_0^1 \big|\widehat{F}_\varrho(r) - F_0(r)\big|\, dr,$$

and perform then a Monte Carlo test. The Monte Carlo test is usually necessary because the distribution of $T_{F_0}(\xi)$ cannot be calculated explicitly, even for relatively simple point processes ξ, like for example a Poisson process. Note that other statistics than T_{F_0} are possible. In fact it is more common to use $\widetilde{T}_{F_0}$ given by $\widetilde{T}_{F_0}(\varrho) = \int_0^1 \big(\widehat{F}_\varrho(r) - F_0(r)\big)^2\, dr$, but T_{F_0} as defined above fits better to our purposes. For more details on the general procedure, see Diggle (1983), Sections 2.2 and 2.4.

Our estimates from Propositions 3.3.F and 3.3.H can now be helpful in several ways. First, it might be that we cannot compute $T_{F_0}(\varrho)$ for a certain point process ξ, because its empty space function F_0 cannot be calculated explicitly, but a bound for $d_W(F_0, F_0')$ is known (e.g. obtained by Proposition 3.3.F) for some other point process η with an empty space function F_0' that *can* be calculated explicitly. We then can estimate the discrepancy between $\widehat{F}_\varrho$ and F_0 as

$$T_{F_0}(\varrho) \leq T_{F_0'}(\varrho) + d_W(F_0, F_0').$$

Secondly, it might be that we would like to simulate from the distribution of $T_{F_0}(\xi)$ in order to perform the Monte Carlo test mentioned above, but that this is too difficult. For one thing the distribution of ξ might either not be known explicitly or at least computationally expensive to simulate from. It might also be that we want to test for a non-simple null hypothesis, for example that ξ has a strong mixing property, so that there is not a fixed distribution for $T_{F_0}(\xi)$ under the null hypothesis, but a whole range. In all of these situations, an upper bound on $d_W\big(\mathscr{L}(T_{F_0}(\xi)), \mathscr{L}(T_{F_0'}(\eta))\big)$, like the one

given in the proposition below, for a point process η whose discrepancy statistic $T_{F_0'}(\eta)$ has a distribution that is easy to simulate from (or even known), is very helpful. We can then compare $T_{F_0}(\varrho)$ with a certain number of values simulated from $\mathscr{L}(T_{F_0'}(\eta))$ rather than $\mathscr{L}(T_{F_0}(\xi))$, and Proposition 3.3.I enables us to adjust the significance level of our test very closely provided that the bound on $d_W\big(\mathscr{L}(T_{F_0}(\xi)), \mathscr{L}(T_{F_0'}(\eta))\big)$ is small. For a detailed description of this procedure, see the example in Subsection 5.3.3, where the same strategy is applied for a slightly altered average nearest neighbor distance statistic $\tilde{U}$ instead of the F-function discrepancy T_{F_0}.

Proposition 3.3.I. *Let W be a compact subset of $\mathbb{R}^D$ that contains $\mathbb{B}(0, 1+\varepsilon)$ for some $\varepsilon > 0$. Let furthermore ξ and η be simple stationary point processes on $\mathbb{R}^D$ that are restricted to W. Then there are constants $c_1, c_2 > 0$ which can be computed explicitly (see Estimate* (3.31) *at the end of the proof), such that*

$$d_W\Big(\mathscr{L}\big(T_{F_0}(\xi)\big), \mathscr{L}\big(T_{F_0'}(\eta)\big)\Big) \le c_1 m d_2\big(\mathscr{L}(\xi), \mathscr{L}(\eta)\big) + c_2 \mathbb{E}\big(|\eta| 1_{\{|\eta|>m\}}\big)$$

for any $m \in \mathbb{N}$.

Proof. We have

$$d_W\Big(\mathscr{L}\big(T_{F_0}(\xi)\big), \mathscr{L}\big(T_{F_0'}(\eta)\big)\Big) = \min_{\substack{X\sim\mathscr{L}(T_{F_0}(\xi))\\ Y\sim\mathscr{L}(T_{F_0'}(\eta))}} \mathbb{E}\big|X-Y\big|.$$

Let $\tilde{\xi}$ and $\tilde{\eta}$ be arbitrary point processes on W with $\tilde{\xi} \stackrel{\mathscr{D}}{=} \xi$ and $\tilde{\eta} \stackrel{\mathscr{D}}{=} \eta$, and set $X := T_{F_0}(\tilde{\xi})$ and $Y := T_{F_0'}(\tilde{\eta})$. Then

$$\begin{aligned}\mathbb{E}\big|X-Y\big| &\le \mathbb{E}\bigg(\int_0^1 \Big|\big|\widehat{F}_{\tilde{\xi}}(r) - F_0(r)\big| - \big|\widehat{F}_{\tilde{\eta}}(r) - F_0'(r)\big|\Big|\, dr\bigg)\\ &\le \mathbb{E}\bigg(\int_0^1 \big|\widehat{F}_{\tilde{\xi}}(r) - \widehat{F}_{\tilde{\eta}}(r)\big|\, dr\bigg) + \int_0^1 \big|F_0(r) - F_0'(r)\big|\, dr. \qquad (3.30)\end{aligned}$$

The second term is just $d_W(F_0, F_0')$, for which we have Remark 3.3.G in combination with Propostion 3.2.C. The first term can be estimated with the aid of Proposition 3.3.H as

$$\begin{aligned}&\mathbb{E}\bigg(\int_0^1 |\widehat{F}_{\tilde{\xi}}(r) - \widehat{F}_{\tilde{\eta}}(r)|\, dr\bigg)\\ &\quad\le \mathbb{E}\bigg(\bigg[|\tilde{\eta}| \int_0^1 \frac{1}{|W^{[-r]}|}\big|\hat{f}_r(\tilde{\xi}) - \hat{f}_r(\tilde{\eta})\big|\, dr\bigg] 1_{\{|\tilde{\xi}|=|\tilde{\eta}|\}}\bigg) + \mathbb{E}\big(d_1(\tilde{\xi},\tilde{\eta}) 1_{\{|\tilde{\xi}|\neq|\tilde{\eta}|\}}\big)\\ &\quad\le \frac{2\alpha_{D-1}}{D|W^{[-1]}|}\mathbb{E}\big(|\tilde{\eta}| d_1(\tilde{\xi},\tilde{\eta}) 1_{\{|\tilde{\xi}|=|\tilde{\eta}|\}}\big) + \mathbb{E}\big(d_1(\tilde{\xi},\tilde{\eta}) 1_{\{|\tilde{\xi}|\neq|\tilde{\eta}|\}}\big)\\ &\quad\le \bigg(\frac{2\alpha_{D-1}}{D|W^{[-1]}|} m \vee 1\bigg)\mathbb{E} d_1(\tilde{\xi},\tilde{\eta}) + \frac{2\alpha_{D-1}}{D|W^{[-1]}|}\mathbb{E}\big(|\tilde{\eta}| 1_{\{|\tilde{\eta}|>m\}}\big)\end{aligned}$$

for any $m \in \mathbb{N}$. In total, we obtain in Inequality (3.30), minimizing over all couplings $(\tilde{\xi}, \tilde{\eta})$ with $\tilde{\xi} \stackrel{\mathscr{D}}{=} \xi$ and $\tilde{\eta} \stackrel{\mathscr{D}}{=} \eta$, that

$$\begin{aligned} &d_W\Big(\mathscr{L}\big(T_{F_0}(\xi)\big), \mathscr{L}\big(T_{F_0'}(\eta)\big)\Big) \\ &\quad \leq \left(\left(\frac{2\alpha_{D-1}}{D|W^{[-1]}|} \vee \frac{1}{m}\right) + 1\right) m d_2\big(\mathscr{L}(\xi), \mathscr{L}(\eta)\big) + \left(\frac{2\alpha_{D-1}}{D|W^{[-1]}|} + 1\right) \mathbb{E}\big(|\eta| 1_{\{|\eta| > m\}}\big), \end{aligned} \tag{3.31}$$

which yields the required result. □

Chapter 4

Stein's method and Poisson process approximation

4.1 An overview of Stein's method

In 1972, Charles Stein published his ingenious method for the normal approximation of dependent random variables (see also Stein (1986) for a broader coverage of topics). The new concept used for this method was the foundation on which many extraordinary results in distributional approximation were to be built. In Chen (1975) an extension of Stein's method to the Poisson distribution was presented, which has been so successful that it is usually referred to as Stein-Chen method. Barbour (1988) introduced an important new view by suggesting the use of generators of Markov processes to obtain the central identity in Stein's method, the so-called Stein equation (see Equation (4.1) below). This made it possible to further extend the method to Poisson process approximation (Barbour (1988) and Barbour and Brown (1992)). Further variants of Stein's method, either derived by Barbour's generator approach or in other ways (see for example the size-bias approach, used among others by Goldstein and Rinott (1996)), include approximation by the compound Poisson distribution (initiated by Barbour, Chen and Loh (1992), see Barbour and Chryssaphinou (2001) for an overview), the compound Poisson process distribution (Barbour and Månsson (2002)), the Wiener measure (Barbour (1990)), the Binomial distribution (Ehm (1991)), the Dirac measure (Reinert (1994)), and the Gamma distribution (Luk (1994), see also Reinert (2005), Section 3).

In 2003, a meeting was held in honor of Charles Stein at the National University of Singapore. The tutorial lectures given during one week of this meeting are collected in the book *An introduction to Stein's method* (edited by A.D. Barbour and L.H.Y. Chen, World Scientific, 2005). Many different ideas, and approaches to Stein's method in most of the above mentioned variants, are treated in detail in this book, and the wealth of results and applications from the different fields makes it not only a sound introductory book for students, but also an excellent reference work for researchers.

A brief sketch of the general idea behind Stein's method could be given as folllows. Suppose that we would like to approximate a complicated probability distribution P on some space E by a simpler probability distribution Q on E in terms of a probability

metric d. We assume here that d is of the form

$$d(P_1, P_2) = \sup_{f \in \mathfrak{F}} \left| \int_E f \, dP_1 - \int_E f \, dP_2 \right|$$

for certain probability measures P_1 and P_2 on E and a set $\mathfrak{F}$ of integrable functions $f : E \to \mathbb{R}$, which is true of many probability metrics (e.g. the total variation metric and the Wasserstein metric, introduced in Section 2.4). The central idea of Stein's method is to re-express the functions $f - \int_E f \, dQ$ as

$$f - \int_E f \, dQ = \mathscr{A}h \tag{4.1}$$

for suitable functions $h = h_f$, where $\mathscr{A}$ is an operator from a subset of $\mathbb{R}^E$ to $\mathbb{R}^E$ that characterizes the distribution Q with which we would like to approximate. Characterizing means here that we have $\int_E (\mathscr{A}h) \, dQ = 0$ for a large enough class of functions h. Equation (4.1) is usually referred to as a *Stein equation* for the given approximation problem. Inserting a random variable X with distribution P and taking expectations yields

$$\left| \int_E f \, dP - \int_E f \, dQ \right| = \left| \mathbb{E}\mathscr{A}h(X) \right|.$$

So what has to be done for bounding $d(P, Q)$ is estimating $\left|\mathbb{E}\mathscr{A}h_f(X)\right|$ uniformly in $f \in \mathfrak{F}$, which surprisingly often turns out to be easier and more successful than bounding the left hand side directly.

The estimation of the right hand side typically involves the computation of upper bounds for the supremum norm of the first and second differences $\|\Delta_1 h\|_\infty$ and $\|\Delta_2 h\|_\infty$, or of the first and second derivatives $\|h'\|_\infty$ and $\|h''\|_\infty$, or of similar quantities, depending on the state space E. For functions h living on $E = \mathbb{Z}_+$, for example, we define the first difference $\Delta h := \Delta_1 h$ by $(\Delta h)(j) := h(j+1) - h(j)$ for every $j \in \mathbb{Z}_+$ and the second difference by $\Delta_2 h := \Delta(\Delta h)$; for h living on $\mathfrak{N}'$ the differences are suitably adapted (see Equations (4.7) and (4.8)). Obtaining bounds for such quantities is often the hardest part of the task. Note that these estimates have nothing to do with the specific random variable X, so once suitable estimates are obtained, they are typically of use for the Q-approximation of many distributions P. The same holds true for the operator $\mathscr{A}$, with the proviso that sometimes there might be several promising operators for a given approximating distribution, and which to choose may then depend on the concrete approximation problem. For some of the most famous examples, however, like the normal, Poisson or Poisson process distribution, there is just one such operator that is considered.

The generator approach by Barbour (1988) suggests choosing for $\mathscr{A}$ the generator of a Markov process on E with stationary distribution Q. This yields for the normal distribution

$$(\mathscr{A}h)(x) = h''(x) - xh'(x) \quad \text{for } x \in \mathbb{R}, \tag{4.2}$$

which defines the generator of the standard Ornstein-Uhlenbeck process (the stationary distribution is the standard normal). For the Poisson distribution, we obtain

$$(\mathscr{A}h)(j) = \lambda\big(h(j+1) - h(j)\big) + j\big(h(j-1) - h(j)\big) \quad \text{for } j \in \mathbb{Z}_+, \tag{4.3}$$

which defines the generator of the immigration-death process with immigration rate $\lambda > 0$ and unit per capita death rate (the stationary distribution is $\mathrm{Po}(\lambda)$). Setting $g(x) := h'(x)$ for all $x \in \mathbb{R}$ in Equation (4.2) yields the famous Stein equation for the normal case, namely

$$f(x) - \mathbb{E}f(Z) = g'(x) - xg(x) \quad \text{for } x \in \mathbb{R},$$

where Z has the standard normal distribution; and setting $g(j) := h(j) - h(j-1)$ for all $j \in \mathbb{N}$ (and assuming e.g. that $g(0) = 0$) in (4.3) yields the famous Stein-Chen equation for the Poisson case, namely

$$f(j) - \mathbb{E}f(Z) = \lambda g(j+1) - jg(j) \quad \text{for } j \in \mathbb{Z}_+,$$

where Z has the $\mathrm{Po}(\lambda)$ distribution. More details on Stein's method in these two cases can be found in Stein (1986), especially Lectures II and III, and Chen and Shao (2005) for the normal case, and in Barbour, Holst and Janson (1992), especially Chapters 1 and 2, and Erhardsson (2005) for the Poisson case.

In what follows, we are of course particularly interested in Stein's method for Poisson process approximation, since Poisson process approximation is the main topic of this thesis. We give a sketch of the key ideas of the method in this case in Section 4.2 and present a collection of the results that are for our purposes the most important ones in Section 4.3. Sometimes we need to have estimates for probability distances between distributions of point counts, and in Section 4.4 we therefore give two results derived by the Stein-Chen method for Poisson approximation.

4.2 Sketch of Stein's method for Poisson process approximation

We rather tersely describe Stein's method for Poisson process approximation in the case of the Barbour-Brown metric d_2. For a detailed presentation in a more general context see Barbour, Holst, and Janson (1992), Chapter 10, and Xia (2005).

Our goal is to bound $d_2(\mathscr{L}(\xi), \mathrm{Po}(\lambda))$ for a point process ξ and a finite measure λ on the compact metric space E'. According to the general procedure laid down in Section 4.1, the key ingredients are the operator $\mathscr{A}$ and bounds of certain differences for the solution h of the Stein equation. Following Barbour's generator approach, we choose for $\mathscr{A}$ the spatial equivalent of the operator given above for the Poisson case. Thus, let $\mathscr{A}$ be the generator of the spatial immigration-death process $(Z(t))_{t\geq 0}$ on E' (that is, the state space is $\mathfrak{N}'$) with immigration measure λ and unit per capita death rate, a pure jump Markov process which has our approximating distribution $\mathrm{Po}(\lambda)$ as equilibrium distribution. This $\mathscr{A}$ is given by

$$\mathscr{A}h(\varrho) = \int_{E'} \big[h(\varrho + \delta_s) - h(\varrho)\big]\ \lambda(ds)\ + \int_{E'} \big[h(\varrho - \delta_s) - h(\varrho)\big]\ \varrho(ds) \tag{4.4}$$

for every $\varrho \in \mathfrak{N}'$ and every $h : \mathfrak{N}' \to \mathbb{R}$ with $\sup_{\varrho:|\varrho|=l} |h(\varrho)| < \infty$ for every $l \in \mathbb{Z}_+$. Note that if the space E' consists just of a single point x and we identify the point measure ϱ on $\{x\}$ with its point count $\varrho(\{x\})$, Equation (4.4) yields the Stein operator for the Poisson distribution from Equation (4.3).

We thus have as Stein equation for the Poisson process case

$$f(\varrho) - \mathbb{E}f(\eta) = \mathscr{A}h(\varrho) \quad \text{for } \varrho \in \mathfrak{N}', \tag{4.5}$$

where $\mathscr{A}$ is as above, $\eta \sim \mathrm{Po}(\lambda)$, and $f \in \mathfrak{F}_2$. Propositions 10.1.1 and 10.1.2 in Barbour, Holst, and Janson (1992) yield that, for each bounded $f : \mathfrak{N}' \to \mathbb{R}$, the function $h := h_f : \mathfrak{N}' \to \mathbb{R}$ that is given by

$$h(\varrho) = h_f(\varrho) = -\int_0^\infty \Big(\mathbb{E}\big(f(Z(t)) \;\big|\; Z(0) = \varrho\big) - \mathbb{E}f(\eta)\Big)\, dt\,, \quad \varrho \in \mathfrak{N}', \tag{4.6}$$

is well-defined, satisfies $\sup_{\varrho:|\varrho|=l}|h(\varrho)| < \infty$ for every $l \in \mathbb{Z}_+$, and solves the Stein equation (4.5). Lemmas 10.2.3 and 10.2.5 of the same book give us the upper bounds we need for the differences of h. We have for any $h = h_f$ with $f \in \mathfrak{F}_2$,

$$\|\Delta_1 h\|_\infty := \sup_{\varrho\in\mathfrak{N},\, s\in E'} \big|h(\varrho + \delta_s) - h(\varrho)\big| \le 1 \wedge \frac{1.65}{\sqrt{|\lambda|}} =: M_1(\lambda), \tag{4.7}$$

and

$$\begin{aligned}\|\Delta_2 h\|_\infty :&= \sup_{\varrho\in\mathfrak{N},\, s_1,s_2\in E'} \big|h(\varrho + \delta_{s_1} + \delta_{s_2}) - h(\varrho + \delta_{s_1}) - h(\varrho + \delta_{s_2}) + h(\varrho)\big| \\ &\le 1 \wedge \left[\frac{2}{|\lambda|}\Big(1 + 2\log^+\Big(\frac{|\lambda|}{2}\Big)\Big)\right] =: M_2(\lambda).\end{aligned} \tag{4.8}$$

As announced in the previous section, Stein's method now suggests bounding $|\mathbb{E}\mathscr{A}h(\xi)|$ uniformly in $f \in \mathfrak{F}_2$, in order to obtain a bound on $d_2(\mathscr{L}(\xi), \mathrm{Po}(\lambda))$.

For the sake of illustration, we demonstrate the short calculation that is necessary for the special case that ξ is a Bernoulli process concentrated on the locations $\alpha_k \in E'$, $k \in \Theta$, where Θ is a finite non-empty index set. This means that $\xi = \sum_{k\in\Theta} I_k\delta_{\alpha_k}$ where the I_k are independent indicators with $\mathbb{E}I_k = q_k \in (0,1)$ for every $k \in \Theta$. We match the expectation measure of the approximated process by requiring that $\lambda = \sum_{k\in\Theta} q_k\delta_k$. Write $\xi_k := \sum_{l\in\Theta, l\ne k} I_l\delta_{\alpha_l}$, and note that $I_k h(\xi - \delta_{\alpha_k}) = I_k h(\xi - I_k\delta_{\alpha_k}) = I_k h(\xi_k)$ and in the same way $I_k h(\xi) = I_k h(\xi_k + I_k\delta_{\alpha_k}) = I_k h(\xi_k + \delta_{\alpha_k})$. Using these two equations for the second relation, and the independence of I_k and ξ_k for third relation, we have

$$\begin{aligned}\big|\mathbb{E}\mathscr{A}h(\xi)\big| &= \left|\mathbb{E}\left(\int_{E'}\big[h(\xi + \delta_s) - h(\xi)\big]\;\lambda(ds) \;+ \int_{E'}\big[h(\xi - \delta_s) - h(\xi)\big]\;\xi(ds)\right)\right| \\ &= \left|\mathbb{E}\left(\sum_{k\in\Theta} q_k\big[h(\xi + \delta_{\alpha_k}) - h(\xi)\big] + \sum_{k\in\Theta} I_k\big[h(\xi_k) - h(\xi_k + \delta_{\alpha_k})\big]\right)\right| \\ &\le \sum_{k\in\Theta} q_k\mathbb{E}\big|h(\xi + \delta_{\alpha_k}) - h(\xi) - h(\xi_k + \delta_{\alpha_k}) + h(\xi_k)\big| \\ &= \sum_{k\in\Theta} q_k^2\mathbb{E}\big|h(\xi_k + 2\delta_{\alpha_k}) - 2h(\xi_k + \delta_{\alpha_k}) + h(\xi_k)\big| \\ &\le M_2(\lambda)\sum_{k\in\Theta} q_k^2.\end{aligned}$$

Thus

$$d_2\big(\mathscr{L}(\xi), \mathrm{Po}(\lambda)\big) = \sup_{f\in\mathfrak{F}_2}\big|\mathbb{E}f(\xi) - \mathbb{E}f(\eta)\big| = \sup_{f\in\mathfrak{F}_2}\big|\mathbb{E}\mathscr{A}h_f(\xi)\big| \le M_2(\lambda)\sum_{k\in\Theta} q_k^2.$$

One of the major strengths of Stein's method is that it usually enables us to treat the case of sums of dependent random variables (here, sums of dependent $I_k\delta_{\alpha_k}$) rather conveniently, whereas other techniques that might work well in the independent case typically break down when it comes to the dependent case. This is true in the above situation as well, and we present one of several possible results for dependent sums $\sum_{k\in\Theta} I_k\delta_{\alpha_k}$ in the next section.

It is furthermore possible to dispose of the $\log^+(|\lambda|/2)$-term entering the above estimate for independent indicators, as is also presented in Section 4.3. Note that for an upper bound on $\|\Delta_2 h\|_\infty$ that is uniform in $f \in \mathfrak{F}_2$, the $\log^+$-term is in general necessary. See the counterexamples in Brown and Xia (1995b).

4.3 Results from Stein's method for Poisson process approximation

There are many different results that are obtained from Stein's method for Poisson process approximation, depending mostly on the state space (continuous or discrete) and the dependence structure of the approximated point process ξ. An additional distinction is the probability metric used, but we have already restricted ourselves to the Barbour-Brown metric in the last section (there are corresponding results with worse bounds for the total variation metric).

In what follows, we always consider a local dependence structure. This means here that $\xi|_{N_\alpha}$, for a not too large neighborhood N_α of α, may depend strongly on $\xi(\{\alpha\})$, but $\xi|_{N_\alpha^c}$ depends only weakly on $\xi(\{\alpha\})$. If the points of ξ are scattered over a continuous space, this has to be re-interpreted in terms of Palm distributions, so that the idea of the weak dependence of $\xi|_{N_\alpha^c}$ on $\xi(\{\alpha\})$ is expressed as $\mathscr{L}(\xi_\alpha|_{N_\alpha^c}) \approx \mathscr{L}(\xi|_{N_\alpha^c})$, where ξ_α is the Palm process of ξ given a point in α (see Kallenberg (1986), Chapter 10, for the definition of the Palm distribution). In any case, it should be clearly noted that the above concept of local dependence serves only for illustration, and does not entail any formal requirements for the corresponding theorems. Formally, the sets N_α are just arbitrary subsets of E that contain α (and satisfy a measurability condition), and they neither have to lie "around α" (they always do in our applications, though), nor to reflect the dependence structure in ξ. However, a judicious choice of the sets N_α is important with reference to the quality of the upper bounds obtained.

We consider local dependence, because this is a rather natural requirement in the approximation problems we have in the following chapters. See for example Barbour, Holst and Janson (1992), Theorem 10.G, for the counterpart of Theorem 4.3.A below that uses a coupling approach for modeling the dependence structure.

Although most of our approximation problems deal with point processes on intervals of $\mathbb{R}^D$, we mainly consider discrete state spaces. This is because the corresponding results from Stein's method are easier to apply, and the necessary step of discretization to get from a point process on an interval in $\mathbb{R}^D$ to a point process on a finite set is usually quite easy and involves only a small error. See for example Barbour and Brown (1992), Theorem 3.6, for the counterpart of Theorem 4.3.A that admits a general state space.

Our processes are indicator point processes on a set $\{\alpha_k;\, k \in \Theta\} \subset E'$, where Θ is a finite non-empty index set. Let $(I_k)_{k\in\Theta}$ be a sequence of indicator random variables

with the local dependence property that for every $k \in \Theta$ the set $\Theta_k := \Theta \setminus \{k\}$ can be partitioned as $\Theta_k = \Theta_k^s \dot\cup \Theta_k^w$ into a set Θ_k^s of indices l for which I_l depends strongly on I_k, and a set Θ_k^w of indices l for which I_l depends weakly on I_k. Again there is no formal requirement for these sets, but it is advisable in view of obtaining good upper bounds to choose them in such a way that they reflect the dependence structure of ξ as well as possible. In the more general notation above, we would have $N_{\alpha_k} = \{\alpha_l;\, l \in \Theta_k^s\} \cup \{\alpha_k\}$ and $N_{\alpha_k}^c = \{\alpha_l;\, l \in \Theta_k^w\}$.

We now write $Z_k := \sum_{l\in\Theta_k^s} I_l$, $\widetilde{Z}_k := Z_k + I_k$, and $q_k := \mathbb{E}I_k \in [0,1]$ for every $k \in \Theta$. Attach the indicators I_k to the locations α_k by setting $\Xi := \sum_{k\in\Theta} I_k\delta_{\alpha_k}$, and define $\lambda := \sum_{k\in\Theta} q_k\delta_{\alpha_k}$, so that $|\lambda| = \sum_{k\in\Theta} q_k$. To exclude a somewhat inconvenient, but trivial special case, assume that $\lambda \neq 0$.

Theorem 4.3.A (Local Stein theorem for indicator point processes). *With the above definitions, we have*

$$d_2\big(\mathscr{L}(\Xi), \mathrm{Po}(\lambda)\big) \le M_2(\lambda) \sum_{k\in\Theta} \big(q_k\mathbb{E}\widetilde{Z}_k + \mathbb{E}(I_kZ_k)\big) + M_1(\lambda) \sum_{k\in\Theta} e_k,$$

where $M_1(\lambda)$ and $M_2(\lambda)$ are given in Inequalities (4.7) *and* (4.8)*, respectively, and*

$$e_k = \mathbb{E}\Big|\mathbb{E}\big(I_k \;\big|\; (I_l\,;\; l \in \Theta_k^w)\big) - q_k\Big| = 2 \max_{B\in\sigma(I_l\,;\, l\in\Theta_k^w)} \big|\mathrm{cov}(I_k, 1_B)\big|.$$

Proof. This follows from Barbour, Holst and Janson (1992), Theorem 10.F, by setting $\boldsymbol{\lambda} := \boldsymbol{\pi}$ in the notation of that theorem. □

Remark 4.3.B. Note that the upper bound in Theorem 4.3.A depends neither on the points α_k, $k \in \Theta$, nor on the specific choice of the metric d_0, as long as it is bounded by 1.

As noted at the end of Section 4.2, the $\log^+(\lambda/2)$-term that enters the bound via $M_2(\lambda)$ is in general necessary if we want to estimate $\|\Delta_2 h\|_\infty$ uniformly in $f \in \mathfrak{F}_2$. However, Brown, Weinberg and Xia (2000) showed that by calculating non-uniform bounds and then using additional arguments corresponding to the concrete situation, it is sometimes possible to dispose of the $\log^+(\lambda/2)$-term. Improvements of their results can be found in Brown and Xia (2001) and in Xia (2005), Sections 5 and 6. Since the additional arguments can be quite involved and since the logarithm above is negligible for almost all practical purposes, we do not pursue this method here much further. We only give a slight adaptation of Equation (6.3) in Theorem 6.1 of Xia (2005), which is needed for the proof of Proposition 6.2.B, where a logarithmic term would be slightly annoying.

Theorem 4.3.C (Independent indicators). *Suppose that $(I_k)_{k\in\Theta}$ is an independent sequence of indicators, and set $q_k := \mathbb{E}I_k$, $\Xi := \sum_k I_k\delta_{\alpha_k}$, and $\lambda := \sum_k q_k\delta_{\alpha_k}$ as above, so that again $|\lambda| = \sum_k q_k$. Then*

$$d_2\big(\mathcal{L}(\Xi), \mathrm{Po}(\lambda)\big) \le \left(1 \wedge \frac{6.5}{|\lambda|}\right) \sum_{k\in\Theta} q_k^2 \le 6.5 \max_{k\in\Theta} q_k.$$

Proof. Suppose that $|\lambda| > 6.5$. From the proof of Theorem 6.1 in Xia (2005), it is immediately clear that

$$d_2\big(\mathcal{L}(\Xi), \mathrm{Po}(\lambda)\big) \le \left(\frac{3.5}{|\lambda|} + \frac{2.5}{|\lambda| - \max_{l\in\Theta} q_l}\right) \sum_{k\in\Theta} q_k^2.$$

Using $q_l \le 1 < |\lambda|/6.5$ we thus obtain

$$d_2\big(\mathcal{L}(\Xi), \mathrm{Po}(\lambda)\big) \le \frac{6.5}{|\lambda|} \sum_{k\in\Theta} q_k^2 \le \sum_{k\in\Theta} q_k^2.$$

For $|\lambda| \le 6.5$, we simply use Theorem 4.3.A with $\Theta_k^s = \emptyset$ and $\Theta_k^w = \Theta_k$ to obtain

$$d_2\big(\mathcal{L}(\Xi), \mathrm{Po}(\lambda)\big) \le \sum_{k\in\Theta} q_k^2 \le \frac{6.5}{|\lambda|} \sum_{k\in\Theta} q_k^2.$$

□

Finally, we give a result for the Barbour-Brown distance between two Poisson processes on the general state space E'.

Theorem 4.3.D. *For finite measures λ and μ on E', we have*

$$d_2\big(\mathrm{Po}(\lambda), \mathrm{Po}(\mu)\big) \le \big(1 - e^{-\min(|\lambda|,|\mu|)}\big)\, d_W\Big(\frac{\lambda}{|\lambda|}, \frac{\mu}{|\mu|}\Big) + \min\Big(1, \frac{1.65}{\sqrt{|\lambda|}}, \frac{1.65}{\sqrt{|\mu|}}\Big)\Big||\lambda| - |\mu|\Big|,$$

where d_W denotes the Wasserstein metric with respect to d_0. The first summand in the upper bound is to be interpreted as zero if either $|\lambda|$ or $|\mu|$ are zero.

Proof. See Brown and Xia (1995a), Inequality (2.8), and the proof following it. Note that the first d_1 in Inequality (2.8) should be d_2; also, there is no formal difference between the two measures in Inequality (2.8), so they can be exchanged in the upper bound. □

4.4 Results from the Stein-Chen method for Poisson approximation

At various occasions, we would like to compute upper bounds for the total variation distance between the distributions of point counts. If one of the point processes involved is a Poisson process, we require results from one-dimensional Poisson approximation, which can be obtained by the original Stein-Chen method.

Theorem 4.4.A (Local Stein-Chen theorem for sums of indicators). *Let Θ, and Θ_k, Θ_k^w, Θ_k^s, I_k, q_k, Z_k, and $\widetilde{Z}_k$ for any $k \in \Theta$ be the same as for Theorem 4.3.A. Set furthermore $W := \sum_{k\in\Theta} I_k$ and $\lambda := \sum_{k\in\Theta} q_k = \mathbb{E}W$. Then*

$$d_{TV}\big(\mathscr{L}(W), \mathrm{Po}(\lambda)\big) \le \left(1 \wedge \frac{1}{\lambda}\right) \sum_{k\in\Theta} \big(q_k \mathbb{E}\widetilde{Z}_k + \mathbb{E}(I_k Z_k)\big) + \left(1 \wedge \frac{1}{\sqrt{\lambda}}\right) \sum_{k\in\Theta} e_k,$$

where again

$$e_k = \mathbb{E}\Big|\mathbb{E}\big(I_k \mid (I_l\,;\ l \in \Theta_k^w)\big) - q_k\Big| = 2 \max_{B\in\sigma(I_l\,;\ l\in\Theta_k^w)} \big|\mathrm{cov}(I_k, 1_B)\big|.$$

Proof. See Barbour, Holst, and Janson (1992), Theorem 1.A. □

Remark 4.4.B ("Magic factors"). Many other methods for bounding the total variation distance between $\mathscr{L}(W)$ and $\mathrm{Po}(\lambda)$, in ways similar to that given above, do not yield the terms $1/\lambda$ and $1/\sqrt{\lambda}$. This fact, together with the general difficulty in fully grasping what friendly forces are at work when Stein's method is applied, forms the reason why the terms $1/\sqrt{\lambda}$ and especially $1/\lambda$ are sometimes called *magic factors* in the context of Stein's method. The same name is used for the corresponding factors $M_2(\lambda)$ and $M_1(\lambda)$ in Theorem 4.3.A and for the factor $1/|\lambda|$ in Theorem 4.3.C.

Still by the original Stein-Chen method, the following upper bound for the distance between two Poisson distributions can be derived.

Theorem 4.4.C. *For any $\lambda, \mu > 0$, we have*

$$d_{TV}\big(\mathrm{Po}(\lambda), \mathrm{Po}(\mu)\big) \leq \min\left(1, \frac{1}{\sqrt{\lambda}}, \frac{1}{\sqrt{\mu}}\right)\big|\lambda - \mu\big|.$$

Proof. This follows in a straightforward manner by applying the Stein-Chen method. Also, it is an immediate consequence of Theorem 1.C in Barbour, Holst, and Janson (1992) if we set $\Lambda \equiv \lambda$. □

A sharper result is proved in Yannaros (1991) by direct calculation of the Hellinger distance between the two Poisson distributions.

Theorem 4.4.D. *For any $\lambda, \mu > 0$, we have*

$$d_{TV}\big(\mathrm{Po}(\lambda), \mathrm{Po}(\mu)\big) \leq \left(1 \wedge \frac{1}{\sqrt{\lambda} + \sqrt{\mu}}\right)\big|\lambda - \mu\big|.$$

Proof. See Yannaros (1991), Theorem 2.1. □

Chapter 5

Approximating linear transforms of point processes

In this chapter, the first of three rather general problems from Poisson process approximation is considered. Using the local Stein Theorem 4.3.A from the last chapter and various results about the Barbour-Brown metric d_2 from Chapter 3, we compute d_2 upper bounds for Poisson process approximations of the linear transforms $\xi\theta_T^{-1}$ that were briefly presented in Subsection 2.2.8. The corresponding convergence theorems were shown in Ellis (1986), and the conditions there are only slightly weaker than what we need for our upper bounds to go to zero. Numerous variants of the bounds are given, and applications to kernel density estimation and long range dependence testing are also presented.

We start out by describing the problem in detail.

5.1 Introduction to the approximation problem

Let $D_1, D_2 \in \mathbb{N}$, $D := D_1 + D_2$, $E := \mathbb{R}^D = \mathbb{R}^{D_1} \times \mathbb{R}^{D_2}$, and let the metric d_0 on E be the Euclidean metric truncated at 1. Consider a point process ξ on E which has expectation measure μ and meets three assumptions, namely absolute continuity of μ with a mild restriction on the density, an orderliness condition in the $\mathbb{R}^{D_1}$-directions and a mixing condition in the $\mathbb{R}^{D_2}$-directions (formal versions of these assumptions can be found at the end of this section). Let η be a Poisson process with the same expectation measure and let $\theta_T : \mathbb{R}^D \to \mathbb{R}^D$ be the linear transformation that stretches the first D_1 coordinates by a factor $w(T)^{1/D_1} > 0$ (typically, $w(T)^{1/D_1} \geq 1$) and compresses the last D_2 coordinates by a factor $T^{1/D_2} \geq 1$, that is for $T \in \mathbb{R}$, $T \geq 1$, we set

$$\theta_T(s,t) := \left(w(T)^{1/D_1} s, \frac{1}{T^{1/D_2}} t \right) \quad \text{for all } (s,t) \in \mathbb{R}^{D_1} \times \mathbb{R}^{D_2} = \mathbb{R}^D,$$

where $w(T) \to \infty$ and $w(T) = O(T)$ for $T \to \infty$. In particular, we usually write $\tilde{\theta}_T$ instead of θ_T if our stretch factor is T^{1/D_1}.

Most of the time we will restrict our transformed processes $\xi\theta_T^{-1}$ and $\eta\theta_T^{-1}$ to the compact cube $E' := J := [-1,1]^D$ and denote by $J_T := \theta_T^{-1}(J)$ the pre-image of J, but sometimes the bigger cuboids $\tilde{J}_T := \tilde{\theta}_T(J_T) = \left[-\left(\frac{T}{w(T)}\right)^{1/D_1}, \left(\frac{T}{w(T)}\right)^{1/D_1}\right]^{D_1} \times [-1,1]^{D_2}$ instead of J are more useful.

A consequence of Theorem 2.5 in Ellis (1986) is that, for bounded measurable functions $f_{1,T}, \ldots, f_{l,T} : J \to \mathbb{R}$ with $\|f_{i,T}\|_\infty = O\big(\sqrt{w(T)/T}\big)$, the distributions of the vectors $\big(\int_J f_{1,T}\, d(\xi\theta_T^{-1}), \ldots, \int_J f_{l,T}\, d(\xi\theta_T^{-1})\big)$ and $\big(\int_J f_{1,T}\, d(\eta\theta_T^{-1}), \ldots, \int_J f_{l,T}\, d(\eta\theta_T^{-1})\big)$ get more and more alike as $T \to \infty$; or more precisely that the difference between their characteristic functions converges to zero as $T \to \infty$. Since the above vectors include (up to a factor of order $\frac{w(T)}{T}$) the fidi-distributions of $\xi\theta_T^{-1}|_J$ and $\eta\theta_T^{-1}|_J$ as special cases, which in turn are intimately connected with the distributions of $\xi\theta_T^{-1}|_J$ and $\eta\theta_T^{-1}|_J$ themselves (see Propositions 2.2.A and 2.2.B), there is hope that $d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\eta\theta_T^{-1}|_J)\big)$ can be shown to be small for large T.

The principal results are given in Section 5.2. It is the main goal of our endeavors to find upper estimates for the distance $d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\eta\theta_T^{-1}|_J)\big)$ (see Subsection 5.2.2), but explicit upper bounds are also computed for $d_{TV}\big(\mathscr{L}\big(\xi\theta_T^{-1}(J)\big), \mathscr{L}\big(\eta\theta_T^{-1}(J)\big)\big)$ (Subsection 5.2.3), $d_2\big(\mathscr{L}(\xi\tilde{\theta}_T^{-1}|_{\tilde{J}_T}), \mathscr{L}(\eta\tilde{\theta}_T^{-1}|_{\tilde{J}_T})\big)$ (Subsection 5.2.4), and $d_2\big(\mathscr{L}(\xi\tilde{\theta}_T^{-1}|_{\tilde{J}_T}), \mathrm{Po}(\lambda|_{\tilde{J}_T})\big)$ for an appropriate T-independent measure λ on $\mathbb{R}^D$ (Subsection 5.2.5).

In Section 5.3, we present some applications of our results. Most importantly, we calculate an upper bound for the bounded Wasserstein distance between the distribution of a kernel estimate for the density of μ at a certain point and the actual value of the density at that point. Furthermore, we briefly describe an application to testing for long range dependence, and give some guidance on how the bounds obtained in the first two sections of this chapter can be used to derive first tentative results for thinning and superposition approximations.

Apart from the paper of Ellis (1986), which provided the initial motivation for many of the theorems in the next section, stretched point processes have also been investigated in the context of light traffic analysis for queues, and in other, similar topics: see e.g. Borovkov (1996) and the references therein. These authors, however, were interested in the quite different question of finding asymptotic expansions for the expectation of functionals of purely stretched marked point processes, which vanish in the limit on every compact set; our procedure, in contrast, leads to point processes with, essentially, a stable or increasing number of points in every compact set.

We conclude this section by taking a detailed look at the three assumptions for the point process ξ.

Assumption 1 (Absolute continuity of the expectation measure). Let $\nu = \nu_1 \otimes \nu_2$, where $\nu_1 := Leb^{D_1}$ is the Lebesgue measure on $\mathbb{R}^{D_1}$, and either $\nu_2 := Leb^{D_2}$ is the Lebesgue measure on $\mathbb{R}^{D_2}$ or $\nu_2 := \mathcal{H}_0^{D_2}$ is the counting measure on $\mathbb{Z}^{D_2} + \frac{1}{2}\mathbf{1} \subset \mathbb{R}^{D_2}$.
Then we require that $\mu \ll \nu$ with a Radon-Nikodym density h such that $\overline{c} \in \mathbb{R}_+$ exists with

$$\overline{c}_T := \sup_{(s,t)\in J_T} h(s,t) \le \overline{c} \quad \text{for all } T \ge 1.$$

In the same way we choose $\underline{c} \in \mathbb{R}_+$ with

$$\underline{c}_T := \inf_{(s,t)\in J_T} h(s,t) \ge \underline{c} \quad \text{for all } T \ge 1.$$

(For the asymptotic result it is enough of course to assume both statements only for all T bigger than some $T_0 \ge 1$.)

Assumption 2 (Orderliness). There is a continuous function $\breve{\alpha} : \mathbb{R}_+ \to \mathbb{R}_+$ with $\breve{\alpha}(0) = 0$, such that for every rectangle $C := [a_1, b_1) \times [a_2, b_2)$ with $a_1, b_1 \in \mathbb{R}^{D_1}$, $a_1 < b_1$, and $a_2, b_2 \in \mathbb{R}^{D_2}$, $a_2 < b_2$, we have

$$\mathbb{E}\big[\big(\xi(C)\big)^2 1_{\{\xi(C)\geq 2\}}\big] \leq v\breve{\alpha}(v),$$

where

$$v := v(C) = \nu_1\big([a_1, b_1)\big)\nu_2\big([a_2, b_2 + \mathbf{1})\big).$$

For the third assumption, there are different versions that can be considered. According to the type of mixing we are interested in, we write this assumption as $3x$, where $x \in \{\beta, \varrho, \varphi\}$.

Assumption 3x (x-mixing property). For every interval $[a_1, b_1) \subset \mathbb{R}^{D_1}$, $a_1 < b_1$, there is a decreasing function $\breve{\beta} := \breve{\beta}_{a_1,b_1} : \mathbb{R}_+ \to \mathbb{R}_+$ with the two following properties:

(i) $\breve{\beta}(u) = o\big(\frac{1}{u^{D_2/2}}\big)$ for $u \to \infty$

(ii) If $a_2, b_2 \in \mathbb{R}^{D_2}$ with $a_2 < b_2$, $t \in \mathbb{R}_+$ and the σ-algebras $\mathcal{F}_{\text{int}}$ and $\mathcal{F}_{\text{ext}}$ are defined as $\mathcal{F}_{\text{int}} := \sigma(\xi|_{[a_1,b_1)\times[a_2,b_2)})$ and $\mathcal{F}_{\text{ext}} := \sigma(\xi|_{[a_1,b_1)\times[a_2-t\mathbf{1},b_2+t\mathbf{1})^c})$, then

$$x(\mathcal{F}_{\text{int}}, \mathcal{F}_{\text{ext}}) \leq \breve{\beta}(t),$$

where x is one of the three mixing coefficients β, ϱ or φ with

$$\begin{aligned}
\beta(\mathcal{F}_{\text{int}}, \mathcal{F}_{\text{ext}}) &:= \mathbb{E}\Big(\operatorname*{ess\,sup}_{B\in\mathcal{F}_{\text{ext}}} \big|\mathbb{P}(B \,|\, \mathcal{F}_{\text{int}}) - \mathbb{P}(B)\big|\Big),\\
\varrho(\mathcal{F}_{\text{int}}, \mathcal{F}_{\text{ext}}) &:= \sup_{\substack{X\in L_2(\mathcal{F}_{\text{int}})\\ Y\in L_2(\mathcal{F}_{\text{ext}})}} \big|\operatorname{corr}(X, Y)\big|,\\
\varphi(\mathcal{F}_{\text{int}}, \mathcal{F}_{\text{ext}}) &:= \sup_{\substack{A\in\mathcal{F}_{\text{int}}\\ B\in\mathcal{F}_{\text{ext}}}} \big|\mathbb{P}(B \,|\, A) - \mathbb{P}(B)\big|.
\end{aligned}$$

For the definition of the essential supremum of an arbitrary set of random variables see Neveu (1965), Proposition II.4.1.

In what follows, we suppress the indication of the interval $[a_1, b_1)$ and write simply $\breve{\beta}$. The corner points a_1 and b_1 are to be chosen appropriately; e.g. $a_1 = -\sup_{T\geq 1}\big(\frac{1}{w(T)}\big)^{1/D_1} \cdot \mathbf{1}$, $b_1 = \sup_{T\geq 1}\big(\frac{1}{w(T)}\big)^{1/D_1} \cdot \mathbf{1}$ is always an appropriate choice.

No further explanation is needed for the first assumption. It simply states the absolute continuity of the expectation measure with respect to what is basically Lebesgue measure, with a mild condition on the density. The fact that we admit the counting measure for the D_2-part of the reference measure ν allows us to apply our future estimates to (mixing) sequences of certain $\mathbb{R}^{D_1}$-valued point processes. In order to simplify certain formulas, we will always tacitly assume that $T \in \{n^{D_2}; n \in \mathbb{N}\}$ if ν_2 is the counting measure.

The second assumption is a form of orderliness in the $\mathbb{R}^{D_1}$-directions. For a detailed account of orderliness see Daley (1974). For what we are interested in here, it is enough to understand that the upper bound for $\mathbb{E}\big[\big(\xi(C)\big)^2 1_{\{\xi(C)\geq 2\}}\big]$ implies that

$$4\,\mathbb{P}[\xi(C) \geq 2] \leq v\breve{\alpha}(v),$$

and that Assumption 2 implies the simplicity of ξ (see the definition in Subsection 2.2.1). The latter implication is due to Theorem 2.6 in Kallenberg (1986).

The various versions of the third assumption are mixing conditions of different strength. It can be seen (Doukhan (1994), Section 1.1) that

$$\begin{aligned} \beta(\mathcal{B},\mathcal{C}) &\leq \varphi(\mathcal{B},\mathcal{C}) \\ \varrho(\mathcal{B},\mathcal{C}) &\leq 2\varphi^{1/2}(\mathcal{B},\mathcal{C})\varphi^{1/2}(\mathcal{C},\mathcal{B}) \end{aligned}$$

for arbitrary σ-algebras $\mathcal{B}$, $\mathcal{C} \subset \mathcal{A}$ on some common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Thus the concept of φ-mixing is the strongest of the three, followed by the β- and ϱ-mixing concepts, which are not generally comparable with each other, although from an empirical point of view β-mixing often turns out to be the stronger of the two. Two mixing concepts that are not treated here are α-mixing, which would be weaker, and ψ-mixing, which would be stronger than any of the three mentioned concepts (see Doukhan (1994)). The kind of mixing used in Ellis (1986) is ϱ-mixing. However, it is important to notice that we need a stronger mixing assumption, in the sense that the set underlying the σ-algebra $\mathcal{F}_{\text{ext}}$ may enclose the set underlying the σ-algebra $\mathcal{F}_{\text{int}}$ from all of the $2D_2$ possible directions of the $\mathbb{R}^{D_2}$. As partial compensation, the order that we need for the convergence of our mixing coefficient to zero is only half that needed for Ellis's result, and what is more, we could actually manage with a mixing assumption where the σ-algebras $\mathcal{F}_{\text{ext}}$ and $\mathcal{F}_{\text{int}}$ are quite a bit smaller (namely generated by the numbers of points of ξ in the corresponding discretization cuboids that we will need for the proof).

5.2 The main results

The results given within this section have somewhat similar flavor, and their proofs all follow the same path; first, discretizing the point processes, and then applying a local Stein theorem. An outline of this method can be found in Subsection 5.2.1; thereafter, in Subsections 5.2.2 to 5.2.5 the different results are presented. A detailed, self-contained proof is given only for Theorem 5.2.A; for the other statements, the necessary adaptations are given.

5.2.1 The approach for solving the approximation problem

All statements in Section 5.2 are about upper bounds for distances between the distribution of a transformed ξ-process and the distribution of a transformed Poisson process (or a function of the respective process, as in Subsection 5.2.3). For the sake of clarity of presentation, we formulate the ideas of the proof only for $d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\eta\theta_T^{-1}|_J)\big)$. However, except for the obvious changes in notation (like writing $\xi\tilde{\theta}_T^{-1}|_{\tilde{J}_T}$ instead of $\xi\theta_T^{-1}|_J$ in Subsection 5.2.4), the arguments presented here can be applied literally (or almost literally in the case of Subsection 5.2.3) to calculate the upper bounds presented for any of the distances appearing in this section.

As mentioned before, our basic strategy of proof is to discretize $\xi\theta_T^{-1}$ and $\eta\theta_T^{-1}$ (in general, the point processes involved) and then apply an estimate, obtained by the local Stein Theorem 4.3.A, to the discretized point processes (in fact, the local Stein-Chen Theorem 4.4.A is enough for Section 5.2.3, where only the numbers of points are involved).

The discretizations are carried out as follows. For every $T \geq 1$ and for $\varsigma(T) \geq 1$ set $n_1 := \lceil \varsigma(T)^{1/D_1} \rceil - 1$ and $n_2 := \lceil T^{1/D_2} \rceil - 1$, where $\lceil x \rceil$ denotes for any $x \in \mathbb{R}$ the smallest integer $z \geq x$. Let $J' := [-1,1)^D$ and $J'_T := \theta_T^{-1}(J')$. Note that $\xi(J_T \setminus J'_T) = \eta(J_T \setminus J'_T) = 0$ almost surely for any T by Assumption 1 (remember that $T \in \{n^{D_2}; n \in \mathbb{N}\}$ if ν_2 is the counting measure), and it is henceforward assumed that $\xi(J_T \setminus J'_T) = \eta(J_T \setminus J'_T) = 0$ on the whole probability space, this having no influence on the distributions of ξ and η. We subdivide J'_T into smaller "discretization cuboids" $C_{\boldsymbol{kl}}$ with lengths 1 in the $\mathbb{R}^{D_2}$-directions and widths $\frac{1}{(w(T)\varsigma(T))^{1/D_1}}$ in the $\mathbb{R}^{D_1}$-directions, whenever the $C_{\boldsymbol{kl}}$ are not too close to the boundary of J'_T. Here $\varsigma(T)$ can be thought of as order of the number of discretization cuboids in the $\mathbb{R}^{D_1}$-directions (there are $2\lceil \varsigma(T)^{1/D_1} \rceil$ in every dimension of $\mathbb{R}^{D_1}$). To be more precise, we set for every $T \geq 1$

$$\begin{aligned} C_{\boldsymbol{kl}} &:= C_{\boldsymbol{kl}}^{(T)} \\ &:= \Bigg(\prod_{r=1}^{D_1} \Big[-\tfrac{n_1}{(w(T)\varsigma(T))^{1/D_1}} + \tfrac{k_r - 1}{(w(T)\varsigma(T))^{1/D_1}}, -\tfrac{n_1}{(w(T)\varsigma(T))^{1/D_1}} + \tfrac{k_r}{(w(T)\varsigma(T))^{1/D_1}} \Big) \\ &\qquad \times \prod_{s=1}^{D_2} \Big[-n_2 + (l_s - 1), -n_2 + l_s \Big) \Bigg) \cap J'_T \end{aligned}$$

$$\text{for all } \boldsymbol{k} = (k_1, k_2, \dots k_{D_1}) \in \{0, 1, \dots, 2n_1 + 1\}^{D_1}$$
$$\text{and } \boldsymbol{l} = (l_1, l_2, \dots l_{D_2}) \in \{0, 1, \dots, 2n_2 + 1\}^{D_2},$$

so that $J'_T = \dot{\bigcup}_{\boldsymbol{k},\boldsymbol{l}} C_{\boldsymbol{kl}}^{(T)}$. Note that in order to reduce the complexity of presentation, we will make use of simplified notations for multi-indices that should be obvious in their meaning. For instance, we write in short $\sum_{\boldsymbol{k}=0}^{2n_1+1} a_{\boldsymbol{k}}$ instead of $\sum_{\boldsymbol{k}:k_1,\dots k_r=0}^{2n_1+1} a_{\boldsymbol{k}}$ or $\boldsymbol{k} \in \{0, 1, \dots, 2n_1 + 1\}$ instead of $\boldsymbol{k} \in \{0, 1, \dots, 2n_1 + 1\}^{D_1}$. Also, where not stated otherwise, the ranges of the indices in expressions like $\sum_{\boldsymbol{k},\boldsymbol{l}}$ or $\bigcup_{\boldsymbol{k},\boldsymbol{l}}$ are given by $\boldsymbol{k} \in \{0, 1, \dots, 2n_1+1\}$, $\boldsymbol{l} \in \{0, 1, \dots, 2n_2+1\}$. Some more notation is needed. We denote by $\alpha_{\boldsymbol{kl}}$ the center of $C_{\boldsymbol{kl}}$, define in the image space of the transformation θ_T

$$\begin{aligned} R_{\boldsymbol{kl}} := R_{\boldsymbol{kl}}^{(T)} := \theta_T(C_{\boldsymbol{kl}}^{(T)}) = \prod_{r=1}^{D_1} \Big[-\tfrac{n_1}{\varsigma(T)^{1/D_1}} + \tfrac{k_r - 1}{\varsigma(T)^{1/D_1}}, -\tfrac{n_1}{\varsigma(T)^{1/D_1}} + \tfrac{k_r}{\varsigma(T)^{1/D_1}} \Big) \\ \times \prod_{s=1}^{D_2} \Big[-\tfrac{n_2}{T^{1/D_2}} + \tfrac{l_s - 1}{T^{1/D_2}}, -\tfrac{n_2}{T^{1/D_2}} + \tfrac{l_s}{T^{1/D_2}} \Big) \end{aligned}$$

for all $\boldsymbol{k}, \boldsymbol{l}$, and write $\gamma_{\boldsymbol{kl}}$ for the center of $R_{\boldsymbol{kl}}$ (correspondingly, we use $\tilde{R}_{\boldsymbol{kl}} := \tilde{\theta}_T(C_{\boldsymbol{kl}}^{(T)})$ and $\tilde{\gamma}_{\boldsymbol{kl}}$ in Subsection 5.2.4).

The discretization Ξ of the point process ξ is obtained by setting a point in the middle of every discretization cuboid $C_{\boldsymbol{kl}}$ which contains any points of ξ. Formally we set

$$I_{\boldsymbol{kl}} := I_{\boldsymbol{kl}}^{(T)} := 1_{\{\xi(C_{\boldsymbol{kl}}) \geq 1\}}, \quad q_{\boldsymbol{kl}} := \mathbb{E} I_{\boldsymbol{kl}} \quad \text{for all } \boldsymbol{k}, \boldsymbol{l},$$
$$W := W^{(T)} := \sum_{\boldsymbol{k},\boldsymbol{l}} I_{\boldsymbol{kl}},$$

and define

$$\Xi := \sum_{\boldsymbol{k},\boldsymbol{l}} I_{\boldsymbol{kl}} \delta_{\alpha_{\boldsymbol{kl}}} \quad \text{and } \tilde{\mu} := \sum_{\boldsymbol{k},\boldsymbol{l}} q_{\boldsymbol{kl}} \delta_{\alpha_{\boldsymbol{kl}}}.$$

Note that $|\tilde{\mu}| = \mathbb{E}W$. The error we make in the transition from $\xi\theta_T^{-1}|_J$ to $\Xi\theta_T^{-1}$ in terms of the d_2-distance (with a slight alteration, the argument holds also for the d_{TV}-distance between the numbers of points; see Subsection 5.2.3) is small for large T, because on the one hand the orderliness condition (Assumption 2) ensures that the probability of two points within the same discretization cuboid (and, as a consequence, of any point vanishing in the transition) is small, and, on the other hand, we have chosen our discretization in such a way that we only have to move points by a d_0-distance of at most half a body diagonal of a discretization cuboid $R_{\boldsymbol{kl}}$ ($\tilde{R}_{\boldsymbol{kl}}$ in Subsection 5.2.4) in the image space, which is small for large T as well.

As a discretization (at least "in distribution") of the Poisson point process η, we take

$$\mathrm{H} := \sum_{\boldsymbol{k},\boldsymbol{l}} Y_{\boldsymbol{kl}} \delta_{\alpha_{\boldsymbol{kl}}},$$

where $Y_{\boldsymbol{kl}}$ are arbitrary independent Poisson distributed random variables with means $q_{\boldsymbol{kl}}$ for $0 \leq \boldsymbol{k} \leq 2n_1 + 1$, $0 \leq \boldsymbol{l} \leq 2n_2 + 1$. Again, the error we make in the transition from $\eta\theta_T^{-1}|_J$ to $\mathrm{H}\theta_T^{-1}$ is small, for reasons quite similar to those stated above for the transition from $\xi\theta_T^{-1}|_J$ to $\Xi\theta_T^{-1}$ (note that the two discretizations are not realized in the same way, and that we have to argue a little more carefully in Subsection 5.2.5, where a limiting Poisson process that is independent of T is considered).

We then have an indicator point process Ξ with a local dependence property (stemming from the mixing Assumption $3x$) and a discrete Poisson point process with the appropriate intensity measure, so that we are in the position to apply the local Stein Theorem 4.3.A for point processes (or, in case of Subsection 5.2.3, Theorem 4.4.A for sums of indicators), which in each case yields the stated result.

There is one point about the refinement of our discretization that is worth noting. In our main ϱ-mixing case we retain the highest possible flexibility by introducing the variable $\varsigma(T)$. Although it will often turn out to be a natural and relatively good choice to set $\varsigma(T) := T$, doing so is in many cases not optimal. The optimal choice of $\varsigma(T)$ depends on the specific orderliness and mixing conditions that can be obtained for ξ. The weaker the orderliness condition (the slower $\breve{\alpha}(v)$ goes to zero for $v \to 0$), the higher the optimal $\varsigma(T)$ will be; conversely (and somewhat surprisingly at the moment), the weaker the mixing condition (the slower $\breve{\beta}(u)$ goes to zero for $u \to \infty$), the lower the optimal $\varsigma(T)$ will be. In contrast, no such considerations are necessary for the discretization in the $\mathbb{R}^{D_2}$-directions. A discretization cuboid length of 1 can easily be seen to be both natural and optimal. A length of higher order in T only increases the distance, by which we have to move points for discretizing, a length of lower order in T increases the number of discretization cuboids without changing the order of the length that the orderliness condition "sees" (i.e. without changing the order of $v(C_{\boldsymbol{kl}})$ with v as in Assumption 2).

5.2.2 The d_2-distance between the point process laws

In this subsection, the d_2-distance between the laws of the transformed point processes $\xi\theta_T^{-1}|_J$ and $\eta\theta_T^{-1}|_J$ is considered. In all the results we use the notation $O\big(f_1(T), \ldots, f_j(T)\big)$ as short hand for $O\big(\max\big\{f_1(T), \ldots, f_j(T)\big\}\big)$.

Results

Theorem 5.2.A (Principal theorem for the transformations $\boldsymbol{\theta_T}$). *Suppose that the prerequisites of Section 5.1 hold, including the Assumptions 1, 2, and 3ϱ, with $\underline{c} > 0$. Then we obtain for arbitrary $m := m(T) \in \mathbb{Z}_+$ and $\varsigma(T) \geq 1$ for every $T \geq 1$:*

$$\begin{aligned} &d_2\Big(\mathscr{L}\big(\xi\theta_T^{-1}|_J\big), \mathscr{L}\big(\eta\theta_T^{-1}|_J\big)\Big) \\ &\quad = O\Big(\frac{1}{\varsigma(T)^{1/D_1}}\,,\ \frac{1}{T^{1/D_2}}\,,\ \log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\frac{m^{D_2}+1}{w(T)}\,, \\ &\qquad \frac{T}{w(T)}\breve{\alpha}\Big(\frac{2^{D_2}}{w(T)\varsigma(T)}\Big)\,,\ \log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\breve{\alpha}\Big(\frac{2^D(2m+1)^{D_2}}{w(T)}\Big)\,,\ \sqrt{T\varsigma(T)}\breve{\beta}(m)\Big) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{for } T \to \infty, \end{aligned}$$

where we write $\log^{\uparrow}(x) := 1 + \big(\log(x) \vee 0\big)$ for $x > 0$.

For a quantitative form of the upper bound see Inequalities (5.10) and (5.11) at the end of the proof. Note that the powers of 2 and 5 that appear in these inequalities have been chosen (for the convenience of calculations) to be unnecessarily large and might be dramatically improved.

One now might ask under what conditions the d_2-distance converges to zero.

Corollary 5.2.B (Convergence to zero in Theorem 5.2.A). *Suppose that the prerequisites of Theorem 5.2.A hold. Furthermore, suppose that $w(T) \geq kT^{\delta}$ for $k > 0$, $\delta \in (0, 1]$ and that*

$$\begin{aligned} &\breve{\alpha}(v) = O\big(v^r\big) \quad \textit{for } v \to 0 && \textit{with } r > 0, \\ &\breve{\beta}(u) = O\Big(\frac{1}{u^{(1+s)D_2/2}}\Big) \quad \textit{for } u \to \infty && \textit{with } 1 + s > \max\Big(\frac{1-\delta}{\delta}\frac{1+r}{r}\,,\ \frac{1}{\delta}\Big). \end{aligned}$$

Then

$$d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\eta\theta_T^{-1}|_J)\big) \longrightarrow 0 \qquad \textit{for } T \to \infty.$$

Remark 5.2.C (Convergence to zero, simplified).

(i) By adjusting m and $\varsigma(T)$ to the function $\breve{\beta}$ it can be shown easily that for $w(T) \asymp T$ the convergence $d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\eta\theta_T^{-1}|_J)\big) \longrightarrow 0$ holds under the general prerequisits of Theorem 5.2.A. This is consistent with Corollary 5.2.B for $\delta = 1$ (note that the requirements for the functions $\breve{\alpha}$ and $\breve{\beta}$ are a bit stronger in 5.2.B).

(ii) From Corollary 5.2.B follows that, for arbitrary $\delta \in (0, 1]$, and for $r > \frac{1-\delta}{1+\delta}$ and $1 + s > \frac{2}{\delta}$, we have $d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\eta\theta_T^{-1}|_J)\big) \longrightarrow 0$ for $T \to \infty$. These simpler, but stronger requirements on the functions $\breve{\alpha}$ and $\breve{\beta}$ reflect the case where we refrain from adapting $\varsigma(T)$ to the concrete problem and simply set $\varsigma(T) = T$.

In the principal Theorem 5.2.A it may seem a little unsatisfactory that our "discretization depth" $\varsigma(T)$ in the $\mathbb{R}^{D_1}$-directions appears in the term $\sqrt{T\varsigma(T)}\breve{\beta}(m)$, which stems from the mixing assumption in the $\mathbb{R}^{D_2}$-directions, and that in fact a finer discretization could increase the overall upper bound we get for the d_2-distance. Whereas it might well

be that the factor $\sqrt{\varsigma(T)}$ is superfluous, it has not been possible to prove this so far. However, there are other ways in which this problem can be, if not remedied, then at least circumvented, simply by assuming one of the other two mixing conditions.

Theorem 5.2.D (Other types of mixing). *Suppose that the requirements for Theorem 5.2.A are met, with the exception that Assumption 3x holds in place of Assumption 3ϱ.*

(i) If x is β, then $d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\eta\theta_T^{-1}|_J)\big)$ has the same order as that stated in Theorem 5.2.A, except for the term $\sqrt{T\varsigma(T)}\breve{\beta}(m)$, which is replaced by the two terms $\sqrt{T/w(T)}\breve{\alpha}\big(2^D/w(T)\big)$ and $\sqrt{w(T)T}\breve{\beta}(m)$; hence (since $\varsigma(T) \geq 1$ was arbitrary)

$$d_2\Big(\mathscr{L}\big(\xi\theta_T^{-1}|_J\big), \mathscr{L}\big(\eta\theta_T^{-1}|_J\big)\Big) = O\bigg(\frac{1}{T^{1/D_2}}\,,\ \log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\frac{m^{D_2}+1}{w(T)}\,,\ \log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\breve{\alpha}\Big(\frac{2^D(2m+1)^{D_2}}{w(T)}\Big)\,,\ \sqrt{\frac{T}{w(T)}}\breve{\alpha}\Big(\frac{2^D}{w(T)}\Big)\,,\ \sqrt{w(T)T}\breve{\beta}(m)\bigg)$$

for $T \to \infty$.

(ii) If x is φ, then $d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\eta\theta_T^{-1}|_J)\big)$ has the same order as that stated in Theorem 5.2.A, but the term $\sqrt{T\varsigma(T)}\breve{\beta}(m)$ can be replaced by $\sqrt{T/w(T)}\breve{\beta}(m)$; hence, as above

$$d_2\Big(\mathscr{L}\big(\xi\theta_T^{-1}|_J\big), \mathscr{L}\big(\eta\theta_T^{-1}|_J\big)\Big) = O\bigg(\frac{1}{T^{1/D_2}}\,,\ \log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\frac{m^{D_2}+1}{w(T)}\,,\ \log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\breve{\alpha}\Big(\frac{2^D(2m+1)^{D_2}}{w(T)}\Big)\,,\ \sqrt{\frac{T}{w(T)}}\breve{\beta}(m)\bigg)$$

for $T \to \infty$.

Remark 5.2.E. Note that in the above theorem a certain price must be paid for the elimination of $\varsigma(T)$ in the term that comes from the mixing condition: In statement (i) we obtain for our upper bound an order which is in many cases worse than the corresponding order we get for an optimal choice of $\varsigma(T)$ in Theorem 5.2.A; only for sufficiently high D_1 is the upper bound order from Theorem 5.2.D(i) in general better. In statement (ii) we require a much stronger kind of mixing condition than in Theorem 5.2.A and Theorem 5.2.D(i).
On the other hand, we do not have to require a strictly stronger mixing condition in statement (i) and we get a strictly better upper bound in statement (ii).

Example. A typical choice of parameters for illustrating the above mentioned points is given by $\breve{\alpha}(v) = v$, $\breve{\beta}(u) = \frac{1}{u^{2D_2}}$, and $w(T) = T$, whence we immediately get $O\big(T^{-1/3}\big)$ and $O\big(T^{-2/3}\big)$ as upper bound orders for the d_2-distance under the β- and φ-mixing assumptions, respectively; solving a little optimization problem yields the order $O\big(T^{-3/(D_1+6)}\big)$ under the ϱ-mixing assumption, which for $D_1 < 3$ is better and for $D_1 > 3$ is worse than the order under β-mixing.

Proofs

The following simple lemma will be useful.

Lemma 5.2.F. *For all $\boldsymbol{k}, \boldsymbol{l}$ we have*

$$\mu(C_{\boldsymbol{kl}}) - 2^{D_2-2}\frac{1}{w(T)\varsigma(T)}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big) \le q_{\boldsymbol{kl}} \le \mu(C_{\boldsymbol{kl}}).$$

Proof. The second inequality is immediate, the first one is obtained as

$$\begin{aligned}
\mu(C_{\boldsymbol{kl}}) - q_{\boldsymbol{kl}} &= \mathbb{E}\xi(C_{\boldsymbol{kl}}) - \mathbb{P}[\xi(C_{\boldsymbol{kl}}) \ge 1] \\
&= \sum_{r=2}^{\infty}(r-1)\mathbb{P}[\xi(C_{\boldsymbol{kl}}) = r] \\
&\le \frac{1}{4}\mathbb{E}\big[\big(\xi(C_{\boldsymbol{kl}})\big)^2 1_{\{\xi(C_{\boldsymbol{kl}})\ge 2\}}\big] \\
&\le 2^{D_2-2}\frac{1}{w(T)\varsigma(T)}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big)
\end{aligned}$$

by the orderliness assumption with $v(C_{\boldsymbol{kl}}) \le 2^{D_2}\frac{1}{w(T)\varsigma(T)}$. □

Proof of Theorem 5.2.A. We use the notation introduced in Subsection 5.2.1; in particular, we write

$$\Xi := \sum_{\boldsymbol{k},\boldsymbol{l}} I_{\boldsymbol{kl}}\delta_{\alpha_{\boldsymbol{kl}}} \quad \text{and} \quad \mathrm{H} := \sum_{\boldsymbol{k},\boldsymbol{l}} Y_{\boldsymbol{kl}}\delta_{\alpha_{\boldsymbol{kl}}}$$

for the discretized point processes, where $Y_{\boldsymbol{kl}}$ are arbitrary independent $\mathrm{Po}(q_{\boldsymbol{kl}})$-variables for $0 \le \boldsymbol{k} \le 2n_1+1$, $0 \le \boldsymbol{l} \le 2n_2+1$.

The overall d_2-distance can now be split up accordingly:

$$\begin{aligned}
d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\eta\theta_T^{-1}|_J)\big) \le\ & d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\Xi\theta_T^{-1})\big) \\
&+ d_2\big(\mathscr{L}(\Xi\theta_T^{-1}), \mathscr{L}(\mathrm{H}\theta_T^{-1})\big) \\
&+ d_2\big(\mathscr{L}(\mathrm{H}\theta_T^{-1}), \mathscr{L}(\eta\theta_T^{-1}|_J)\big). \qquad (5.1)
\end{aligned}$$

We first take a look at the discretization errors. For the ξ-discretization we can obtain via the Kantorovich-Rubinstein Theorem 3.2.A(i) for d_2 (see also Inequality (3.7)) that

$$\begin{aligned}
d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\Xi\theta_T^{-1})\big) &\le \mathbb{E}d_1(\xi\theta_T^{-1}|_J, \Xi\theta_T^{-1}) \\
&= \mathbb{E}\big(d_1(\xi\theta_T^{-1}|_J, \Xi\theta_T^{-1})1_{\{\xi\theta_T^{-1}(J)=W\}}\big) \\
&\quad + 1\cdot\mathbb{P}[\xi\theta_T^{-1}(J) \ne W]. \qquad (5.2)
\end{aligned}$$

The second summand can easily be estimated as follows.

$$\begin{aligned}
\mathbb{P}[\xi\theta_T^{-1}(J) \ne W] &= \mathbb{P}\big[\textstyle\bigcup_{\boldsymbol{k},\boldsymbol{l}}\{\xi(C_{\boldsymbol{kl}}) \ge 2\}\big] \\
&\le \sum_{\boldsymbol{k},\boldsymbol{l}}\mathbb{P}[\xi(C_{\boldsymbol{kl}}) \ge 2] \\
&\le \frac{1}{4}\sum_{\boldsymbol{k},\boldsymbol{l}}\mathbb{E}\big[\big(\xi(C_{\boldsymbol{kl}})\big)^2 1_{\{\xi(C_{\boldsymbol{kl}})\ge 2\}}\big] \\
&\le 2^{2D+D_2-2}\frac{T}{w(T)}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big) \qquad (5.3)
\end{aligned}$$

by the orderliness assumption with $v(C_{\boldsymbol{kl}}) \le 2^{D_2}\frac{1}{w(T)\varsigma(T)}$.

In order to estimate the first summand in (5.2), we use the representation of the d_1-distance given in Lemma 3.1.A. Let $S_1, \dots, S_{\xi\theta_T^{-1}(J)}$ be the points of $\xi\theta_T^{-1}|_J$ and $S'_1, \dots, S'_W$ the points of $\Xi\theta_T^{-1}$ and suppose w.l.o.g. that they are numbered in an optimal way on $\{\xi\theta_T^{-1}(J) = W\}$, that is, in such a way that S'_i is the center $\gamma_{\boldsymbol{kl}}$ of the cuboid $R_{\boldsymbol{kl}}$ which contains S_i. Thus by Lemma 3.1.A, and since in the transition from ξ to Ξ we do not move the points any farther than half a body diagonal of a cuboid $R_{\boldsymbol{kl}}$,

$$\begin{aligned} &d_1(\xi\theta_T^{-1}|_J, \Xi\theta_T^{-1})1_{\{\xi\theta_T^{-1}(J)=W\}} \\ &\qquad = \Big(\frac{1}{W}\sum_{i=1}^{W} d_0(S_i, S'_i)\Big)1_{\{\xi\theta_T^{-1}(J)=W\ge 1\}} \\ &\qquad \le \frac{1}{2}\sqrt{D_1\Big(\frac{1}{\varsigma(T)^{1/D_1}}\Big)^2 + D_2\Big(\frac{1}{T^{1/D_2}}\Big)^2}\,1_{\{\xi\theta_T^{-1}(J)=W\ge 1\}} \\ &\qquad \le \frac{1}{2}\Big(\frac{\sqrt{D_1}}{\varsigma(T)^{1/D_1}} + \frac{\sqrt{D_2}}{T^{1/D_2}}\Big), \end{aligned} \tag{5.4}$$

whence we get for the total ξ-discretization error

$$\begin{aligned} &d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\Xi\theta_T^{-1})\big) \\ &\qquad \le \frac{1}{2}\Big(\frac{\sqrt{D_1}}{\varsigma(T)^{1/D_1}} + \frac{\sqrt{D_2}}{T^{1/D_2}}\Big) + 2^{2D+D_2-2}\frac{T}{w(T)}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big). \end{aligned}$$

Next we consider the discretization error for η. Let $\mathrm{H}' := \sum_{\boldsymbol{k},\boldsymbol{l}} \eta(C_{\boldsymbol{kl}})\delta_{\alpha_{\boldsymbol{kl}}}$ and $q'_{\boldsymbol{kl}} := \mu(C_{\boldsymbol{kl}})$. We split up the error as

$$\begin{aligned} &d_2\big(\mathscr{L}(\mathrm{H}\theta_T^{-1}), \mathscr{L}(\eta\theta_T^{-1}|_J)\big) \\ &\qquad \le d_2\big(\mathscr{L}(\mathrm{H}\theta_T^{-1}), \mathscr{L}(\mathrm{H}'\theta_T^{-1})\big) + d_2\big(\mathscr{L}(\mathrm{H}'\theta_T^{-1}), \mathscr{L}(\eta\theta_T^{-1}|_J)\big). \end{aligned} \tag{5.5}$$

The first summand gives us a little more trouble. Since d_2 is a bounded Wasserstein metric because of $d_1 \le 1$, and $d_{BW} \le d_{TV}$ is generally true by Figure 2.4.1 and the corresponding proof, it follows that $d_2 \le d_{TV}$. By application of Theorem 2.4.A for getting the second and third lines, we then have

$$\begin{aligned} d_2\big(\mathscr{L}(\mathrm{H}\theta_T^{-1}), \mathscr{L}(\mathrm{H}'\theta_T^{-1})\big) &\le d_{TV}\big(\mathscr{L}(\mathrm{H}\theta_T^{-1}), \mathscr{L}(\mathrm{H}'\theta_T^{-1})\big) \\ &\le \min_{\substack{Y_{\boldsymbol{kl}}\sim\mathrm{Po}(q_{\boldsymbol{kl}}),\perp\!\!\!\perp \\ Y'_{\boldsymbol{kl}}\sim\mathrm{Po}(q'_{\boldsymbol{kl}}),\perp\!\!\!\perp}} \sum_{\boldsymbol{k},\boldsymbol{l}} \mathbb{P}\big[Y_{\boldsymbol{kl}} \ne Y'_{\boldsymbol{kl}}\big] \\ &= \sum_{\boldsymbol{k},\boldsymbol{l}} d_{TV}\big(\mathrm{Po}(q_{\boldsymbol{kl}}), \mathrm{Po}(q'_{\boldsymbol{kl}})\big) \\ &\le \sum_{\boldsymbol{k},\boldsymbol{l}} \big(q'_{\boldsymbol{kl}} - q_{\boldsymbol{kl}}\big) \\ &\le 2^{2D+D_2-2}\frac{T}{w(T)}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big), \end{aligned} \tag{5.6}$$

where the last two inequalities follow from Theorem 4.4.D and Lemma 5.2.F, respectively. For the second summand in (5.5) we obtain

$$\begin{aligned} d_2\big(\mathscr{L}(\mathrm{H}'\theta_T^{-1}), \mathscr{L}(\eta\theta_T^{-1}|_J)\big) &\le \mathbb{E}d_1(\mathrm{H}'\theta_T^{-1}, \eta\theta_T^{-1}|_J) \\ &= \mathbb{E}\big[d_1(\mathrm{H}'\theta_T^{-1}, \eta\theta_T^{-1}|_J)1_{\{\mathrm{H}'\theta_T^{-1}(J)=\eta\theta_T^{-1}(J)\}}\big] \\ &\le \frac{1}{2}\bigg(\frac{\sqrt{D_1}}{\varsigma(T)^{1/D_1}} + \frac{\sqrt{D_2}}{T^{1/D_2}}\bigg) \end{aligned} \tag{5.7}$$

by the same argument that was used in Inequality (5.4). So, an estimate for the total η-discretization error is given by

$$\begin{aligned} &d_2\big(\mathscr{L}(\mathrm{H}\theta_T^{-1}), \mathscr{L}(\eta\theta_T^{-1}|_J)\big) \\ &\qquad \le \frac{1}{2}\bigg(\frac{\sqrt{D_1}}{\varsigma(T)^{1/D_1}} + \frac{\sqrt{D_2}}{T^{1/D_2}}\bigg) + 2^{2D+D_2-2}\frac{T}{w(T)}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big). \end{aligned}$$

Last, we look at the remaining term $d_2\big(\mathscr{L}(\Xi\theta_T^{-1}), \mathscr{L}(\mathrm{H}\theta_T^{-1})\big)$, which is perfect for the application of a Stein estimate. In the notation of Section 4.3 we write

$$\Theta = \{0, 1, \dots, 2n_1 + 1\}^{D_1} \times \{0, 1, \dots, 2n_2 + 1\}^{D_2}$$

(accordingly we write the elements of Θ as $(\boldsymbol{i}, \boldsymbol{j})$, meaning $\boldsymbol{i} \in \{0, 1, \dots, 2n_1 + 1\}^{D_1}$, $\boldsymbol{j} \in \{0, 1, \dots, 2n_2 + 1\}^{D_2}$), and for the sets of strongly and weakly dependent indicators, respectively,

$$\begin{aligned} \Theta^s_{\boldsymbol{kl}} &= \{(\boldsymbol{i}, \boldsymbol{j}) \in \Theta_{\boldsymbol{kl}}\,;\ |\boldsymbol{j} - \boldsymbol{l}| \le m\} \ \text{ and} \\ \Theta^w_{\boldsymbol{kl}} &= \{(\boldsymbol{i}, \boldsymbol{j}) \in \Theta_{\boldsymbol{kl}}\,;\ |\boldsymbol{j} - \boldsymbol{l}| \ge m + 1\} \end{aligned}$$

for every $\boldsymbol{k}$, $\boldsymbol{l}$, where $|\boldsymbol{j} - \boldsymbol{l}| := \max_{1\le s\le D_2}|j_s - l_s|$ and $m := m(T) \in \mathbb{Z}_+$ for every $T \ge 1$ is chosen arbitrarily. We can assume without loss of generality that $m \le 2n_2 + 1$ (note that for $m > 2n_2 + 1$ we have $e_{\boldsymbol{kl}} = 0$, so that Inequality (5.9) below is still true). Also as in Section 4.3, we set

$$Z_{\boldsymbol{kl}} := \sum_{(\boldsymbol{i},\boldsymbol{j})\in\Theta^s_{\boldsymbol{kl}}} I_{\boldsymbol{ij}}.$$

From the local Stein Theorem 4.3.A for point processes we know that

$$\begin{aligned} &d_2\big(\mathscr{L}(\Xi\theta_T^{-1}), \mathscr{L}(\mathrm{H}\theta_T^{-1})\big) \\ &\qquad \le \bigg\{1 \wedge \bigg[\frac{2}{|\tilde{\mu}|}\Big(1 + 2\log^+\Big(\frac{|\tilde{\mu}|}{2}\Big)\Big)\bigg]\bigg\}\sum_{\boldsymbol{k},\boldsymbol{l}}\big(q^2_{\boldsymbol{kl}} + q_{\boldsymbol{kl}}\mathbb{E}Z_{\boldsymbol{kl}} + \mathbb{E}(I_{\boldsymbol{kl}}Z_{\boldsymbol{kl}})\big) \\ &\qquad\quad + \bigg(1 \wedge \frac{1.65}{\sqrt{|\tilde{\mu}|}}\bigg)\sum_{\boldsymbol{k},\boldsymbol{l}} e_{\boldsymbol{kl}} \end{aligned} \tag{5.8}$$

with

$$e_{\boldsymbol{kl}} = 2\max_{B\in\sigma(I_{\boldsymbol{ij}}\,;\,(\boldsymbol{i},\boldsymbol{j})\in\Theta^w_{\boldsymbol{kl}})}\big|\mathrm{cov}(I_{\boldsymbol{kl}}, 1_B)\big|.$$

Starting from the right hand side, most further estimates are very easy. First, we have

$$q_{\boldsymbol{kl}} \le \mu(C_{\boldsymbol{kl}}) \le \overline{c}_T \frac{1}{w(T)\varsigma(T)}, \quad |\tilde{\mu}| = \sum_{\boldsymbol{k},\boldsymbol{l}} q_{\boldsymbol{kl}} \le \mu(J_T) \le 2^D \overline{c}_T \frac{T}{w(T)},$$

and

$$\mathbb{E}Z_{\boldsymbol{kl}} = \sum_{\boldsymbol{i}=0}^{2n_1+1} \sum_{\substack{\boldsymbol{j}=(\boldsymbol{l}-m)\vee 0 \\ (\boldsymbol{i},\boldsymbol{j})\neq(\boldsymbol{k},\boldsymbol{l})}}^{(\boldsymbol{l}+m)\wedge(2n_2+1)} q_{\boldsymbol{ij}} \le \overline{c}_T \big[(2n_1+2)^{D_1}(2m+1)^{D_2} - 1\big] \frac{1}{w(T)\varsigma(T)};$$

furthermore, by the mixing assumption,

$$\begin{aligned} e_{\boldsymbol{kl}} = 2\sqrt{q_{\boldsymbol{kl}}(1-q_{\boldsymbol{kl}})} \max_{B\in\sigma\left(I_{\boldsymbol{ij}}\,;\,(\boldsymbol{i},\boldsymbol{j})\in\Theta_{\boldsymbol{kl}}^{w}\right)} & \sqrt{\mathbb{P}[B]\big(1-\mathbb{P}[B]\big)}\big|\mathrm{corr}(I_{\boldsymbol{kl}}, 1_B)\big| \\ & \le 2\sqrt{q_{\boldsymbol{kl}}}\frac{1}{2}\breve{\beta}(m) \le \sqrt{\overline{c}_T}\sqrt{\frac{1}{w(T)\varsigma(T)}}\breve{\beta}(m); \end{aligned} \tag{5.9}$$

and, by Lemma 5.2.F,

$$\begin{aligned} |\tilde{\mu}| = \sum_{\boldsymbol{k},\boldsymbol{l}} q_{\boldsymbol{kl}} &\ge \sum_{\boldsymbol{k},\boldsymbol{l}} \left(\mu(C_{\boldsymbol{kl}}) - 2^{D_2-2}\frac{1}{w(T)\varsigma(T)}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big)\right) \vee 0 \\ &= \left(\mu(J_T) \;-\; (2n_1+2)^{D_1}(2n_2+2)^{D_2}\frac{2^{D_2-2}}{w(T)\varsigma(T)}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big)\right) \vee 0 \\ &\ge 2^D \frac{T}{w(T)}\left(\underline{c}_T \;-\; 2^{D+D_2-2}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big)\right) \vee 0, \end{aligned}$$

whence we get a "magic factor" estimate of

$$\frac{1}{|\tilde{\mu}|} \le \big(1+\varepsilon(T)\big)\frac{1}{2^D\underline{c}_T}\frac{w(T)}{T}$$

with

$$\varepsilon(T) := \begin{cases} \left(1 - 2^{D+D_2-2}\frac{1}{\underline{c}_T}\breve{\alpha}\big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\big)\right)^{-1} - 1 & \text{if } \Big(1-\ldots\Big) > 0 \\ \infty & \text{otherwise,} \end{cases}$$

an expression of order $O\big(\breve{\alpha}\big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\big)\big)$ for $T\to\infty$, provided that $\underline{c} > 0$.

For the remaining term, $\mathbb{E}(I_{\boldsymbol{kl}}Z_{\boldsymbol{kl}})$, a little trick is required. We subdivide the set $\Theta = \{0,1,\ldots,2n_1+1\}^{D_1}\times\{0,1,\ldots,2n_2+1\}^{D_2}$ along the last D_2 dimensions in D_2-cube sections of extension $2m+1$ in every dimension (except for possible left over cuboids), and look at the individual sections separately. For $\boldsymbol{r} = (r_1, r_2, \ldots, r_{D_2}) \in \big\{1,2,\ldots,\big\lceil\frac{2n_2+2}{2m+1}\big\rceil\big\}^{D_2}$ set for the $\boldsymbol{r}$-th section, i.e. the section containing the r_j-th collection of $2m+1$ numbers in the j-th coordinate,

$$\begin{aligned} \boldsymbol{z}^{(\mathbf{1})}(\boldsymbol{r}) := \boldsymbol{z}^{(\mathbf{1})}(\boldsymbol{r}, m) &:= \big(z_1^{(1)}(\boldsymbol{r}), \ldots, z_{D_2}^{(1)}(\boldsymbol{r})\big) \\ &:= \big((r_1-1)(2m+1), \ldots, (r_{D_2}-1)(2m+1)\big), \end{aligned}$$

which is the "lower left" corner index (the multi-index that is in each coordinate minimal among all indices belonging to the $\boldsymbol{r}$-th section), and

$$\begin{aligned}\boldsymbol{z^{(2)}}(\boldsymbol{r}) &:= \boldsymbol{z^{(2)}}(\boldsymbol{r},m) := \big(z_1^{(2)}(\boldsymbol{r}),\dots,z_{D_2}^{(2)}(\boldsymbol{r})\big)\\ &:= \big([r_1(2m+1)-1]\wedge(2n_2+1),\dots,[r_{D_2}(2m+1)-1]\wedge(2n_2+1)\big),\end{aligned}$$

which is the "upper right" corner index (the multi-index that is in each coordinate maximal among all indices belonging to the $\boldsymbol{r}$-th section). Furthermore, let

$$M_{\boldsymbol{r}} := M_{\boldsymbol{r}}^{(m)} := \bigcup_{\boldsymbol{i}=0}^{2n_1+1} \bigcup_{\boldsymbol{j}=[\boldsymbol{z^{(1)}}(\boldsymbol{r})-m]\vee 0}^{[\boldsymbol{z^{(2)}}(\boldsymbol{r})+m]\wedge(2n_2+1)} C_{\boldsymbol{ij}}$$

be the subset of J_T that naturally belongs to the m-neighborhood cube of the $\boldsymbol{r}$-th section. Using our usual multi-index notation and index range convention for sums, we now obtain for the remaining term

$$\begin{aligned}&\sum_{\boldsymbol{k},\boldsymbol{l}} \mathbb{E}(I_{\boldsymbol{kl}}Z_{\boldsymbol{kl}})\\
&\quad= \mathbb{E}\Bigg(\sum_{\boldsymbol{k}=0}^{2n_1+1}\sum_{\boldsymbol{l}=0}^{2n_2+1}\sum_{\boldsymbol{i}=0}^{2n_1+1}\sum_{\substack{\boldsymbol{j}=(\boldsymbol{l}-m)\vee 0\\(\boldsymbol{i},\boldsymbol{j})\neq(\boldsymbol{k},\boldsymbol{l})}}^{(\boldsymbol{l}+m)\wedge(2n_2+1)} I_{\boldsymbol{kl}}I_{\boldsymbol{ij}}\Bigg)\\
&\quad\le \mathbb{E}\Bigg\{\sum_{\boldsymbol{r}=1}^{\lceil\frac{2n_2+2}{2m+1}\rceil}\Bigg(\sum_{\boldsymbol{k}=0}^{2n_1+1}\sum_{\boldsymbol{l}=\boldsymbol{z^{(1)}}(\boldsymbol{r})}^{\boldsymbol{z^{(2)}}(\boldsymbol{r})}\sum_{\boldsymbol{i}=0}^{2n_1+1}\sum_{\substack{\boldsymbol{j}=[\boldsymbol{z^{(1)}}(\boldsymbol{r})-m]\vee 0\\(\boldsymbol{i},\boldsymbol{j})\neq(\boldsymbol{k},\boldsymbol{l})}}^{[\boldsymbol{z^{(2)}}(\boldsymbol{r})+m]\wedge(2n_2+1)} I_{\boldsymbol{kl}}I_{\boldsymbol{ij}}\Bigg)1_{\{\xi(M_{\boldsymbol{r}}^{(m)})\ge 2\}}\Bigg\}\\
&\quad\le \mathbb{E}\Bigg\{\sum_{\boldsymbol{r}=1}^{\lceil\frac{2n_2+2}{2m+1}\rceil}\Bigg(\sum_{\boldsymbol{i}=0}^{2n_1+1}\sum_{\boldsymbol{j}=[\boldsymbol{z^{(1)}}(\boldsymbol{r})-m]\vee 0}^{[\boldsymbol{z^{(2)}}(\boldsymbol{r})+m]\wedge(2n_2+1)} I_{\boldsymbol{ij}}\Bigg)^2 1_{\{\xi(M_{\boldsymbol{r}}^{(m)})\ge 2\}}\Bigg\}\\
&\quad\le \sum_{\boldsymbol{r}=1}^{\lceil\frac{2n_2+2}{2m+1}\rceil}\mathbb{E}\Big(\big(\xi(M_{\boldsymbol{r}}^{(m)})\big)^2 1_{\{\xi(M_{\boldsymbol{r}}^{(m)})\ge 2\}}\Big)\\
&\quad\le 2^{D+D_2}\big(T^{1/D_2}+m+1\big)^{D_2}\frac{1}{w(T)}\breve{\alpha}\Big(2^D(2m+1)^{D_2}\frac{1}{w(T)}\Big)\end{aligned}$$

by the orderliness assumption with $v(M_{\boldsymbol{r}}^{(m)}) \le 2^D(2m+1)^{D_2}\frac{1}{w(T)}$.

All that is left to do now is to combine the various estimates for the right hand side terms of Inequality (5.8). Then, adding the discretization errors and setting

$$L(T) := 1 \wedge \left[2\big(1+\varepsilon(T)\big)\frac{w(T)}{2^D\underline{c}_T T}\Big(1+2\log^+\Big(2^{D-1}\overline{c}_T\frac{T}{w(T)}\Big)\Big)\right]$$

yields for the overall d_2-distance

$$\begin{aligned}
&d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J),\mathscr{L}(\eta\theta_T^{-1}|_J)\big)\\
&\qquad\le \frac{\sqrt{D_1}}{\varsigma(T)^{1/D_1}}+\frac{\sqrt{D_2}}{T^{1/D_2}}+L(T)2^{2D+2D_1}\overline{c}_T^{\,2}\frac{T(2m+1)^{D_2}}{(w(T))^2}\\
&\qquad\quad+2^{2D+D_2-1}\frac{T}{w(T)}\breve{\alpha}\Big(\frac{2^{D_2}}{w(T)\varsigma(T)}\Big)\\
&\qquad\quad+L(T)2^{D+D_2}\frac{(T^{1/D_2}+m+1)^{D_2}}{w(T)}\breve{\alpha}\Big(2^D\frac{(2m+1)^{D_2}}{w(T)}\Big)\\
&\qquad\quad+\left(1\wedge 1.65\sqrt{1+\varepsilon(T)}\sqrt{\frac{w(T)}{2^D\underline{c}_T T}}\right)2^{2D}\sqrt{\overline{c}_T}\sqrt{\frac{\varsigma(T)}{w(T)}}T\breve{\beta}(m).
\end{aligned}\tag{5.10}$$

For $\underline{c}>0$ and preferably T large enough we get the rougher, but less nasty looking upper bound

$$\begin{aligned}
&d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J),\mathscr{L}(\eta\theta_T^{-1}|_J)\big)\\
&\qquad\le \frac{\sqrt{D_1}}{\varsigma(T)^{1/D_1}}+\frac{\sqrt{D_2}}{T^{1/D_2}}\\
&\qquad\quad+2^{D+2D_1+2}\frac{\overline{c}^2}{\underline{c}}\big(1+\varepsilon(T)\big)\log^{\uparrow}\Big(2^{D-1}\overline{c}\frac{T}{w(T)}\Big)\frac{(2m+1)^{D_2}}{w(T)}\\
&\qquad\quad+2^{2D+D_2-1}\frac{T}{w(T)}\breve{\alpha}\Big(\frac{2^{D_2}}{w(T)\varsigma(T)}\Big)\\
&\qquad\quad+2^{D_2+2}\,5^{D_2}\,\frac{1}{\underline{c}}\big(1+\varepsilon(T)\big)\log^{\uparrow}\Big(2^{D-1}\overline{c}\frac{T}{w(T)}\Big)\breve{\alpha}\Big(2^D\frac{(2m+1)^{D_2}}{w(T)}\Big)\\
&\qquad\quad+2^{\frac{3}{2}D+1}\sqrt{\frac{\overline{c}}{\underline{c}}}\sqrt{1+\varepsilon(T)}\sqrt{T\varsigma(T)}\breve{\beta}(m),
\end{aligned}\tag{5.11}$$

which is of the required order. □

Proof of Corollary 5.2.B. For $T\ge 1$ we have to find $\varsigma(T)\ge 1$ and $m:=m(T)\in\mathbb{Z}_+$, such that all six terms on the right hand side of the equality in Theorem 5.2.A go to zero as $T\to\infty$. We set $\varsigma(T):=T^y$ and $m:=\big[T^x\big]$, with $y>0$ and $0\le x<\frac{\delta}{D_2}$. Thus

$$\begin{aligned}
\frac{1}{\varsigma(T)^{1/D_1}}\longrightarrow 0,\quad \frac{1}{T^{1/D_2}}\longrightarrow 0,\\
\log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\frac{m^{D_2}+1}{w(T)}\longrightarrow 0,\quad\text{and}\\
\log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\breve{\alpha}\Big(\frac{2^D(2m+1)^{D_2}}{w(T)}\Big)\longrightarrow 0;
\end{aligned}$$

so the only two terms we have to worry about are

$$\frac{T}{w(T)}\breve{\alpha}\Big(\frac{2^{D_2}}{w(T)\varsigma(T)}\Big)=O\big(T^{1-\delta-\delta r-yr}\big)$$

and

$$\sqrt{T\varsigma(T)}\breve{\beta}(m) = O\big(T^{\frac{1}{2}(1+y-(1+s)D_2x)}\big),$$

which both converge to zero if there exist $y > 0$ and $0 \le x < \frac{\delta}{D_2}$ such that

$$y > \frac{1-\delta-\delta r}{r} \quad \text{and} \quad y < (1+s)D_2x - 1.$$

This last is true provided that

$$(1+s)\delta - 1 > \max\left(\frac{1-\delta-\delta r}{r}\,,\,0\right),$$

whence we obtain the statement. □

Proof of Theorem 5.2.D. Since the mixing condition is used only once in the proof of Theorem 5.2.A, namely in Inequality (5.9) for obtaining the upper bound of the $e_{\boldsymbol{kl}}$ from the Stein estimate, we can simply transfer the proof and re-calculate this upper bound under our new mixing conditions.

(i) Let $\boldsymbol{l} \in \{0,1,\dots,2n_2+1\}^{D_2}$ be fixed, set $C_{\cdot\boldsymbol{l}} := \bigcup_{\boldsymbol{k}=0}^{2n_1+1} C_{\boldsymbol{kl}}$, and define

$$\tilde{I}^{(\boldsymbol{l})}_{\text{int}} := \big(I_{\boldsymbol{il}}\,;\ \boldsymbol{i} \in \{0,1,\dots,2n_1+1\}^{D_1}\big), \qquad \tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{int}} := \sigma(\tilde{I}^{(\boldsymbol{l})}_{\text{int}})$$
$$\tilde{I}^{(\boldsymbol{l})}_{\text{ext}} := \big(I_{\boldsymbol{ij}}\,;\ (\boldsymbol{i},\boldsymbol{j}) \in \Theta^{w}_{\boldsymbol{kl}}\big),\ \text{regardless of } \boldsymbol{k}, \qquad \tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{ext}} := \sigma(\tilde{I}^{(\boldsymbol{l})}_{\text{ext}}).$$

Note that $\tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{int}} \subset \mathcal{F}^{(\boldsymbol{l})}_{\text{int}} := \sigma(\xi|_{C_{\cdot\boldsymbol{l}}})$ and $\tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{ext}} \subset \mathcal{F}^{(\boldsymbol{l})}_{\text{ext}} := \sigma(\xi|_{\bigcup_{(\boldsymbol{i},\boldsymbol{j})\in\Theta^{w}_{\boldsymbol{kl}}} C_{\boldsymbol{ij}}})$, regardless of $\boldsymbol{k}$. It is seen for every $\boldsymbol{k} \in \{0,1,\dots,2n_1+1\}^{D_1}$ that

$$\begin{aligned}
e_{\boldsymbol{kl}} &= 2\max_{B\in\tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{ext}}}\big|\text{cov}(I_{\boldsymbol{kl}},1_B)\big| \\
&= 2\max_{B\in\tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{ext}}}\big|\mathbb{P}[B\cap\{I_{\boldsymbol{kl}}=1\}] - \mathbb{P}[B]\mathbb{P}[I_{\boldsymbol{kl}}=1]\big| \\
&\le 2\max_{B\in\tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{ext}}}\Big|\mathbb{P}\big[B\cap\big\{\tilde{I}^{(\boldsymbol{l})}_{\text{int}}=x_{\boldsymbol{k}}\big\}\big] - \mathbb{P}\big[B\big]\mathbb{P}\big[\tilde{I}^{(\boldsymbol{l})}_{\text{int}}=x_{\boldsymbol{k}}\big]\Big| \\
&\quad + 2\max_{B\in\tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{ext}}}\Big|\mathbb{P}\big[B\cap\big\{I_{\boldsymbol{kl}}=1\big\}\cap\big\{\textstyle\sum_{\boldsymbol{i}} I_{\boldsymbol{il}}\ge 2\big\}\big] \\
&\qquad\qquad - \mathbb{P}\big[B\big]\mathbb{P}\big[\big\{I_{\boldsymbol{kl}}=1\big\}\cap\big\{\textstyle\sum_{\boldsymbol{i}} I_{\boldsymbol{il}}\ge 2\big\}\big]\Big|,
\end{aligned}$$

where $x_{\boldsymbol{k}}$ is the element of $\{0,1\}^{\{0,1,\dots,2n_1+1\}^{D_1}}$ which has a 1 in the $\boldsymbol{k}$-th and a 0 in every other component. We denote the first summand by $A_{\boldsymbol{kl}}$, the second by $B_{\boldsymbol{kl}}$ and look at the sums over $\boldsymbol{k}$ separately. For the $A_{\boldsymbol{kl}}$-sum we obtain

$$\begin{aligned}
\sum_{\boldsymbol{k}=0}^{2n_1+1} A_{\boldsymbol{kl}} &= 2\sum_{\boldsymbol{k}=0}^{2n_1+1}\max_{B\in\tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{ext}}}\Big|\mathbb{P}\big[B\mid\tilde{I}^{(\boldsymbol{l})}_{\text{int}}=x_{\boldsymbol{k}}\big]-\mathbb{P}\big[B\big]\Big|\,\mathbb{P}\big[\tilde{I}^{(\boldsymbol{l})}_{\text{int}}=x_{\boldsymbol{k}}\big] \\
&\le 2\,\mathbb{E}\bigg(\max_{B\in\tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{ext}}}\Big|\mathbb{P}\big[B\mid\tilde{I}^{(\boldsymbol{l})}_{\text{int}}\big]-\mathbb{P}\big[B\big]\Big|\bigg) \\
&= 2\,\beta\big(\tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{int}},\tilde{\mathcal{F}}^{(\boldsymbol{l})}_{\text{ext}}\big) \le 2\,\beta\big(\mathcal{F}^{(\boldsymbol{l})}_{\text{int}},\mathcal{F}^{(\boldsymbol{l})}_{\text{ext}}\big) \le 2\,\breve{\beta}(m),
\end{aligned}$$

where the monotony of the β-mixing coefficient is immediate if it is written in its dual form as a supremum over measurable partitions (see Doukhan (1994), Section 1.1). For the $B_{\boldsymbol{kl}}$-sum the upper bound is obtained by application of the orderliness condition:

$$\begin{aligned}\sum_{\boldsymbol{k}=0}^{2n_1+1} B_{\boldsymbol{kl}} &\le 4\sum_{\boldsymbol{k}=0}^{2n_1+1}\mathbb{E}\big(I_{\boldsymbol{kl}}1_{\{\sum_{\boldsymbol{i}} I_{\boldsymbol{il}}\ge 2\}}\big)\\ &\le 2\,\mathbb{E}\big[\big(\xi(C_{\cdot\boldsymbol{l}})\big)^2 1_{\{\xi(C_{\cdot\boldsymbol{l}})\ge 2\}}\big]\\ &\le 2^{D+1}\frac{1}{w(T)}\breve{\alpha}\Big(2^D\frac{1}{w(T)}\Big).\end{aligned}$$

We thus have for the total $e_{\boldsymbol{kl}}$-sum over $\boldsymbol{k}$ the estimate

$$\sum_{\boldsymbol{k}=0}^{2n_1+1} e_{\boldsymbol{kl}} \le 2\breve{\beta}(m) + 2^{D+1}\frac{1}{w(T)}\breve{\alpha}\Big(2^D\frac{1}{w(T)}\Big).$$

(ii) In the case of the φ-mixing condition the corresponding estimate is very easy. It follows that

$$\begin{aligned}e_{\boldsymbol{kl}} &= 2\max_{B\in\tilde{\mathcal{F}}_{\mathrm{ext}}^{(\boldsymbol{l})}}\big|\mathrm{cov}(I_{\boldsymbol{kl}},1_B)\big|\\ &= 2\left(\max_{B\in\tilde{\mathcal{F}}_{\mathrm{ext}}^{(\boldsymbol{l})}}\big|\mathbb{P}[B\,|\,I_{\boldsymbol{kl}}=1]-\mathbb{P}[B]\big|\right)\mathbb{P}[I_{\boldsymbol{kl}}=1]\\ &\le 2\,\breve{\beta}(m)\frac{\overline{c}_T}{w(T)\varsigma(T)}.\end{aligned}$$

□

5.2.3 The d_{TV}-distance between the laws of the point counts

In Inequality (3.15) of Subsection 3.3.1 we have seen that

$$d_{TV}\big(\mathscr{L}\big(\xi_1(E')\big),\mathscr{L}\big(\xi_2(E')\big)\big) \le d_2\big(\mathscr{L}(\xi_1),\mathscr{L}(\xi_2)\big)$$

for any two point processes ξ_1 and ξ_2 on a general compact space E'. Thus the upper bounds we obtained in the theorems of Subsection 5.2.2 are also upper bounds for $d_{TV}\big(\mathscr{L}\big(\xi\theta_T^{-1}(J)\big),\mathscr{L}\big(\eta\theta_T^{-1}(J)\big)\big)$. However, using the same method as above and making only slight modifications in the proofs one can do a little better. Note that although now we are only concerned about numbers of points and not about their positions, we can still improve (but possibly also impair, depending on the leading term in our estimate) our upper bound by choosing a finer discretization in the $\mathbb{R}^{D_1}$-directions. This is because the advantage we get from the orderliness assumption if we have smaller discretization cuboids surmounts the disadvantage of having more of them.

Theorem 5.2.G. *Suppose that the prerequisites of Section 5.1 hold, including the Assumptions 1, 2, and 3ϱ, with $\underline{c} > 0$.*

Then we obtain for arbitrary $m := m(T) \in \mathbb{Z}_+$ and $\varsigma(T) \geq 1$ for every $T \geq 1$:

$$d_{TV}\Big(\mathscr{L}\big(\xi\theta_T^{-1}(J)\big), \mathscr{L}\big(\eta\theta_T^{-1}(J)\big)\Big) = O\bigg(\frac{m^{D_2}+1}{w(T)}\,,\ \frac{T}{w(T)}\breve{\alpha}\Big(\frac{2^{D_2}}{w(T)\varsigma(T)}\Big)\,,\ \breve{\alpha}\Big(\frac{2^D(2m+1)^{D_2}}{w(T)}\Big)\,,\ \sqrt{T\varsigma(T)}\breve{\beta}(m)\bigg) \quad \text{for } T \to \infty.$$

Remark 5.2.H. Of course, all theorems stated in Subsection 5.2.2 have their equivalents for the d_{TV}-distance between the distributions of the total point counts. The corresponding upper bounds can simply be obtained by leaving out the $\log^{\uparrow}$-terms, as well as the terms

$$\frac{1}{\varsigma(T)^{1/D_1}} \quad \text{and} \quad \frac{1}{T^{1/D_2}}.$$

Note, however, that the conditions in Corollary 5.2.B for convergence to zero of the principal upper bound remain unchanged.

Proof of Theorem 5.2.G. Although our task now seems to be quite different, we can proceed exactly as we did in the proof of Theorem 5.2.A. First we split up the distance as

$$\begin{aligned} d_{TV}\big(\mathscr{L}\big(\xi\theta_T^{-1}(J)\big), \mathscr{L}\big(\eta\theta_T^{-1}(J)\big)\big) &= d_{TV}\big(\mathscr{L}\big(\xi(J_T)\big), \mathrm{Po}(\mu(J_T))\big) \\ &\leq d_{TV}\big(\mathscr{L}(\xi(J_T)), \mathscr{L}(W)\big) + d_{TV}\big(\mathscr{L}(W), \mathrm{Po}(|\tilde{\mu}|)\big) \\ &\quad + d_{TV}\big(\mathrm{Po}(|\tilde{\mu}|), \mathrm{Po}(\mu(J_T))\big). \end{aligned}$$

Here, the two discretization errors can be estimated very easily. By the orderliness assumption we obtain

$$\begin{aligned} d_{TV}\big(\mathscr{L}(\xi(J_T)), \mathscr{L}(W)\big) &\leq \mathbb{P}[\xi(J_T) \neq W] \\ &= \mathbb{P}\big[\textstyle\bigcup_{\boldsymbol{k},\boldsymbol{l}}\{\xi(C_{\boldsymbol{kl}}) \geq 2\}\big] \\ &\leq \frac{1}{4}\sum_{\boldsymbol{k},\boldsymbol{l}} \mathbb{E}\big[\big(\xi(C_{\boldsymbol{kl}})\big)^2 1_{\{\xi(C_{\boldsymbol{kl}})\geq 2\}}\big] \\ &\leq 2^{2D+D_2-2}\frac{T}{w(T)}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big) \end{aligned}$$

and by Theorem 4.4.D and Lemma 5.2.F

$$\begin{aligned} d_{TV}\big(\mathrm{Po}(|\tilde{\mu}|), \mathrm{Po}(\mu(J_T))\big) &\leq \bigg(1 \wedge \frac{1}{\sqrt{|\tilde{\mu}|} + \sqrt{\mu(J_T)}}\bigg)\big||\tilde{\mu}| - \mu(J_T)\big| \\ &\leq \bigg(1 \wedge \frac{1}{\sqrt{\mu(J_T)}}\bigg)\sum_{\boldsymbol{k},\boldsymbol{l}}\big(\mu(C_{\boldsymbol{kl}}) - q_{\boldsymbol{kl}}\big) \\ &\leq \bigg(1 \wedge \bigg[\frac{1}{2^{D/2}\sqrt{\underline{c}_T}}\sqrt{\frac{w(T)}{T}}\,\bigg]\bigg)2^{2D+D_2-2}\frac{T}{w(T)}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big). \end{aligned}$$

As for the remaining term, $d_{TV}\big(\mathscr{L}(W), \mathrm{Po}(|\tilde{\mu}|)\big)$, we can proceed exactly as we did with $d_2\big(\mathscr{L}(\Xi\theta_T^{-1}), \mathscr{L}(\mathrm{H}\theta_T^{-1})\big)$, with the only difference that now we use the local Stein-Chen Theorem 4.4.A. Thus

$$\begin{aligned} &d_{TV}\big(\mathscr{L}(W), \mathrm{Po}(|\tilde{\mu}|)\big) \\ &\qquad \le \Big(1 \wedge \frac{1}{|\tilde{\mu}|}\Big) \sum_{\boldsymbol{k},\boldsymbol{l}} \big(q_{\boldsymbol{kl}}^2 + q_{\boldsymbol{kl}} \mathbb{E} Z_{\boldsymbol{kl}} + \mathbb{E}(I_{\boldsymbol{kl}} Z_{\boldsymbol{kl}})\big) + \Big(1 \wedge \frac{1}{\sqrt{|\tilde{\mu}|}}\Big) \sum_{\boldsymbol{k},\boldsymbol{l}} e_{\boldsymbol{kl}} \end{aligned}$$

with

$$e_{\boldsymbol{kl}} = 2 \max_{B \in \sigma(I_{\boldsymbol{ij}}\,;\,(\boldsymbol{i},\boldsymbol{j}) \in \Theta_{\boldsymbol{kl}}^{w})} \big|\mathrm{cov}(I_{\boldsymbol{kl}}, 1_B)\big|.$$

All notation has exactly the same meaning as it had in the proof of Theorem 5.2.A, so except for the logarithmic factor in front of the first sum, and the constant 1.65 in front of the second, we get exactly the same upper bound for $d_{TV}\big(\mathscr{L}(W), \mathrm{Po}(|\tilde{\mu}|)\big)$ as we did for $d_2\big(\mathscr{L}(\Xi\theta_T^{-1}), \mathscr{L}(\mathrm{H}\theta_T^{-1})\big)$.

Assembling of all the different pieces yields the result claimed. □

5.2.4 Results for the measure preserving transformations $\tilde{\boldsymbol{\theta}}_{\boldsymbol{T}}$

When we consider a stretch factor $w(T)^{1/D_1} = o(T^{1/D_1})$, the expected number of points of the transformed process $\xi\theta_T^{-1}$ contained within the fixed cube J goes to infinity as $T \to \infty$ if $\underline{c} > 0$, which for some applications is not desirable (e.g. if we want to approximate $\xi\theta_T^{-1}|_J$ by a Poisson process that does not depend on T, see Subsection 5.2.5). We therefore formulate another theorem in this section, which deals with the case where we adjust the volume of the cuboid J to the volume of the cuboids J_T, and thus produce space for the additional points.

In this regard, let $\tilde{\theta}_T$ and $\tilde{J}_T$, defined as in Section 5.1, be our substitute for the transformation θ_T and our enlarged version of the cuboid J, respectively. We then obtain the following result, where once more the quantitative form of the upper bound can be found at the end of the proof.

Theorem 5.2.I. *Suppose that the prerequisites of Section 5.1 hold, including the Assumptions 1, 2, and 3ϱ, with $\underline{c} > 0$.*
Then we obtain for arbitrary $m := m(T) \in \mathbb{Z}_+$ and $\varsigma(T) \ge 1$ for every $T \ge 1$:

$$\begin{aligned} &d_2\Big(\mathscr{L}\big(\xi\tilde{\theta}_T^{-1}|_{\tilde{J}_T}\big), \mathscr{L}\big(\eta\tilde{\theta}_T^{-1}|_{\tilde{J}_T}\big)\Big) \\ &\quad = O\bigg(\Big(\frac{T}{w(T)}\Big)^{1/D_1} \frac{1}{\varsigma(T)^{1/D_1}}\,,\ \frac{1}{T^{1/D_2}}\,,\ \log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\frac{m^{D_2}+1}{w(T)}\,, \\ &\qquad \frac{T}{w(T)}\breve{\alpha}\Big(\frac{2^{D_2}}{w(T)\varsigma(T)}\Big)\,,\ \log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\breve{\alpha}\Big(\frac{2^D(2m+1)^{D_2}}{w(T)}\Big)\,,\ \sqrt{T\varsigma(T)}\breve{\beta}(m)\bigg) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{for } T \to \infty, \end{aligned}$$

which is the same order as in Theorem 5.2.A apart from the factor $\big(T/w(T)\big)^{1/D_1}$.

Proof. For a large part we can adopt the proof of Theorem 5.2.A. We use the same notation and the same discretization as we did there, replacing only θ_T by $\tilde{\theta}_T$ and J by $\tilde{J}_T$. First note that there is no change at all for the estimate of the Stein term, now written as $d_2\big(\mathscr{L}(\Xi\tilde{\theta}_T^{-1}), \mathscr{L}(\mathrm{H}\tilde{\theta}_T^{-1})\big)$, because in the Stein estimate only objects in the pre-image of $\tilde{\theta}_T$ have to be considered (the Stein estimate does not take into account the distances between the points!).

But the changes for the estimates of the approximation errors are not exactly huge either: As we have seen in the proof of Theorem 5.2.A, these errors can be split up into two additive parts, one stemming from the fact that the original and the discretized point process need not have the same numbers of points in every discretization cuboid (see Inequalities (5.3) resp. (5.6) in the proof of Theorem 5.2.A), and one stemming from the fact that even when we have the same numbers of points in every discretization cuboid, their positions are in general a bit shifted (see Inequalities (5.4) resp. (5.7)). From those two parts only the second is affected by the transition from θ_T to $\tilde{\theta}_T$ and from J to $\tilde{J}_T$ (inasmuch as the discretization cuboids in the image space get a little bigger), because for the first we have to deal once more only with objects in the pre-image of $\tilde{\theta}_T$. A short calculation taking into account the above considerations (reproducing Inequalities (5.4) and accordingly (5.7)) provides as upper bounds for each of the discretization errors $d_2\big(\mathscr{L}(\xi\tilde{\theta}_T^{-1}|_{\tilde{J}_T}), \mathscr{L}(\Xi\tilde{\theta}_T^{-1})\big)$ and $d_2\big(\mathscr{L}(\mathrm{H}\tilde{\theta}_T^{-1}), \mathscr{L}(\eta\tilde{\theta}_T^{-1}|_{\tilde{J}_T})\big)$

$$\frac{1}{2}\bigg(\Big(\frac{T}{w(T)}\Big)^{1/D_1}\frac{\sqrt{D_1}}{\varsigma(T)^{1/D_1}} + \frac{\sqrt{D_2}}{T^{1/D_2}}\bigg) + 2^{2D+D_2-2}\frac{T}{w(T)}\breve{\alpha}\Big(2^{D_2}\frac{1}{w(T)\varsigma(T)}\Big).$$

Thus we obtain as possible upper bounds for the overall d_2-distance those of Inequalities (5.10) and (5.11) with $\frac{\sqrt{D_1}}{\varsigma(T)^{1/D_1}}$ replaced by $\big(\frac{T}{w(T)}\big)^{1/D_1}\frac{\sqrt{D_1}}{\varsigma(T)^{1/D_1}}$, which yields the required qualitative estimate. □

Again we can formulate versions of the other results of Subsection 5.2.2 with only slight (and very obvious) changes; in particular we get

Corollary 5.2.J (Convergence to zero in Theorem 5.2.I). *Suppose that the prerequisites of Theorem 5.2.I hold. Furthermore, suppose that $w(T) \geq kT^\delta$ for $k > 0$, $\delta \in (0, 1]$ and that*

$$\breve{\alpha}(v) = O\big(v^r\big) \quad \text{for } v \to 0 \qquad \text{with } r > 0,$$
$$\breve{\beta}(u) = O\Big(\frac{1}{u^{(1+s)D_2/2}}\Big) \quad \text{for } u \to \infty \qquad \text{with } 1 + s > \max\Big(\frac{1-\delta}{\delta}\frac{1+r}{r}\,,\ \frac{2-\delta}{\delta}\Big).$$

Then

$$d_2\big(\mathscr{L}(\xi\tilde{\theta}_T^{-1}|_{\tilde{J}_T}), \mathscr{L}(\eta\tilde{\theta}_T^{-1}|_{\tilde{J}_T})\big) \longrightarrow 0 \qquad \text{for } T \to \infty.$$

Note that under the β- or the φ-mixing assumption no changes in the respective upper bound order obtained in Theorem 5.2.D are necessary.

5.2.5 Results for a fixed limiting process

So far we have only examined approximations of the transformed process $\xi\theta_T^{-1}$ (respectively $\xi\tilde{\theta}_T^{-1}$) by a Poisson process which has the expectation measure $\mu\theta_T^{-1}$. Of course,

this implies that the expectation measure may (and, unless it is a constant multiple of the Lebesgue measure, does) change as T tends to infinity: the approximating Poisson process in general will not be stable. One might therefore ask, under what circumstances it is possible to approximate the transformed ξ-process by a fixed Poisson process, whose distribution does not depend on T, and what loss in terms of the d_2-distance one has to face.

First of all, the correct T-independent intensity measure for our new Poisson process has to be found. Clearly, for $\underline{c} > 0$ using the transformation θ_T with a stretch factor $w(T) = o(T)$ is unnatural, because in that case the expected number of points of $\xi\theta_T^{-1}$ contained in J goes to infinity, whereas, of course, for any fixed Poisson process the expectation of the number of points in J is always finite. So the natural choice for general $w(T)$ is the measure preserving transformation $\tilde{\theta}_T$ together with the enlarged cuboid $\tilde{J}_T$ from Subsection 5.2.4.

For the following heuristics we ignore the fact that ν_2 might be a counting measure. Then, restricted to the cuboid J_T for T relatively large, the measure μ with density h with respect to Leb^D should be relatively "close" to the measure $\lambda := h(0)Leb^D$, provided that h is constant in the $\mathbb{R}^{D_2}$-directions (hence the notation $h(s) = h(s,t)$ for $s \in \mathbb{R}^{D_1}$, $t \in \mathbb{R}^{D_2}$) and that h satisfies a regularity condition in the $\mathbb{R}^{D_1}$-directions at 0. Thus, restricted to $\tilde{J}_T$, $\mu\tilde{\theta}_T^{-1}$ should be close to $\lambda\tilde{\theta}_T^{-1}$ (which is again $h(0)Leb^D$, hence not dependent on T) as well, and therefore $\mathrm{Po}\big(h(0)Leb^D|_{\tilde{J}_T}\big)$ should be a good choice for approximating $\mathscr{L}(\xi\theta_T^{-1}|_{\tilde{J}_T})$.

The following makes the above considerations rigorous. First we formulate the additional regularity condition for h.

Assumption 4 (Regularity of h). The density $h = d\mu/d\nu$ is constant in the $\mathbb{R}^{D_2}$-directions, so that we can write

$$h(s,t) = h(s) \quad \text{for all } s \in \mathbb{R}^{D_1}, t \in \mathbb{R}^{D_2} \text{ (resp. } t \in \mathbb{Z}^{D_2} + \tfrac{1}{2}\mathbf{1}).$$

Moreover, h satisfies the following regularity condition in the $\mathbb{R}^{D_1}$-directions: There exist $L \geq 0$ and $z > 0$, such that

$$\big|h(s) - h(0)\big| \leq L|s|^z \quad \text{for all } s \in \mathbb{R}^{D_1}$$

(or for $s \in \big[-\big(\frac{1}{w(T)}\big)^{1/D_1}, \big(\frac{1}{w(T)}\big)^{1/D_1}\big)^{D_1}$ for the T one wishes to consider).

We are now in the position to formulate the theorem.

Theorem 5.2.K. *Suppose that the prerequisites of Section 5.1 hold, including the Assumptions 1, 2, 3ϱ, as well as the new Assumption 4 above. Let $\underline{c} > 0$, $T \geq 1$ (remember that we always assume that $T \in \{n^{D_2}; n \in \mathbb{N}\}$ if $\nu_2 = \mathcal{H}_0^{D_2}$), $m := m(T) \in \mathbb{Z}_+$, and $\varsigma(T) \geq 1$. Then*

$$\begin{aligned} d_2\Big(\mathscr{L}\big(\xi\tilde{\theta}_T^{-1}|_{\tilde{J}_T}\big), \mathrm{Po}\big(h(0)Leb^D|_{\tilde{J}_T}\big)\Big) \;&\leq\; \tilde{A}(T) + 2^{D_2} D_1^{\frac{z+D_1}{2}} \frac{D_1}{z+D_1} L\alpha_{D_1} \frac{T}{w(T)^{1+z/D_1}} \\ &= O\Big(\frac{T}{w(T)^{1+z/D_1}}, \Big(\frac{T}{w(T)}\Big)^{1/D_1} \frac{1}{\varsigma(T)^{1/D_1}}, \frac{1}{T^{1/D_2}}, \log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\frac{m^{D_2}+1}{w(T)}, \\ &\qquad \frac{T}{w(T)}\breve{\alpha}\Big(\frac{2^{D_2}}{w(T)\varsigma(T)}\Big), \log^{\uparrow}\Big(\frac{T}{w(T)}\Big)\breve{\alpha}\Big(\frac{2^D(2m+1)^{D_2}}{w(T)}\Big), \sqrt{T\varsigma(T)}\breve{\beta}(m)\Big) \end{aligned}$$

for $T \to \infty$,

where $\tilde{A}(T) := \tilde{A}(T, m, \varsigma(T))$ is the explicit upper bound that we obtained in Theorem 5.2.I (Formula (5.10) *or* (5.11) *with the corresponding modifications) and α_{D_1} denotes the volume of the D_1-dimensional unit ball as usual.*

Corollary 5.2.L. *Under the prerequisites of Corollary 5.2.J plus Assumption 4 with $z > \frac{1-\delta}{\delta} D_1$, we obtain*

$$d_2\big(\mathscr{L}(\xi\tilde{\theta}_T^{-1}|_{\tilde{J}_T}), \mathrm{Po}(h(0)Leb^D|_{\tilde{J}_T})\big) \longrightarrow 0 \qquad \text{for } T \to \infty,$$

and hence, if $\delta = 1$ ($z > 0$),

$$\xi\tilde{\theta}_T^{-1}|_J \xrightarrow{\mathscr{D}} \mathrm{Po}(h(0)Leb^D|_J)$$

by Theorem 3.2.A(ii).

Proof of Theorem 5.2.K. Once again we can largely adopt the proof of Theorem 5.2.A (or more precisely that of Theorem 5.2.I). This time only the estimate for the discretization error $d_2\big(\mathscr{L}(\mathrm{H}\tilde{\theta}_T^{-1}), \mathscr{L}(\eta\tilde{\theta}_T^{-1}|_{\tilde{J}_T})\big)$ has to be replaced by an appropriate estimate for our new error $d_2\big(\mathscr{L}(\mathrm{H}\tilde{\theta}_T^{-1}), \mathrm{Po}(h(0)Leb^D|_{\tilde{J}_T})\big)$. We proceed just as we did in Theorem 5.2.A.

Let $\eta' \sim \mathrm{Po}(h(0)Leb^D)$ (hence, by Proposition 2.2.R, $\eta'\tilde{\theta}_T^{-1} \sim \mathrm{Po}(h(0)Leb^D)$), $\mathrm{H}'' := \sum_{\boldsymbol{k},\boldsymbol{l}} \eta'(C_{\boldsymbol{kl}})\delta_{\alpha_{\boldsymbol{kl}}}$, and split up the error as

$$\begin{aligned} d_2\big(\mathscr{L}(\mathrm{H}\tilde{\theta}_T^{-1}), \mathrm{Po}(h(0)Leb^D|_{\tilde{J}_T})\big) &= d_2\big(\mathscr{L}(\mathrm{H}\tilde{\theta}_T^{-1}), \mathscr{L}(\eta'\tilde{\theta}_T^{-1}|_{\tilde{J}_T})\big) \\ &\le d_2\big(\mathscr{L}(\mathrm{H}\tilde{\theta}_T^{-1}), \mathscr{L}(\mathrm{H}''\tilde{\theta}_T^{-1})\big) + d_2\big(\mathscr{L}(\mathrm{H}''\tilde{\theta}_T^{-1}), \mathscr{L}(\eta'\tilde{\theta}_T^{-1}|_{\tilde{J}_T})\big). \end{aligned}$$

Inequality (5.7) (or more precisely the corresponding modification from the proof of Theorem 5.2.I) yields for the second summand, as before,

$$d_2\big(\mathscr{L}(\mathrm{H}''\tilde{\theta}_T^{-1}), \mathscr{L}(\eta'\tilde{\theta}_T^{-1}|_{\tilde{J}_T})\big) \le \frac{1}{2}\bigg(\Big(\frac{T}{w(T)}\Big)^{1/D_1} \frac{\sqrt{D_1}}{\varsigma(T)^{1/D_1}} + \frac{\sqrt{D_2}}{T^{1/D_2}}\bigg). \tag{5.12}$$

For the first summand we get by the same method as in Inequality (5.6)

$$\begin{aligned} d_2\big(\mathscr{L}(\mathrm{H}\tilde{\theta}_T^{-1}), \mathscr{L}(\mathrm{H}''\tilde{\theta}_T^{-1})\big) &\le \sum_{\boldsymbol{k},\boldsymbol{l}} d_{TV}\big(\mathrm{Po}(q_{\boldsymbol{kl}}), \mathrm{Po}(h(0)Leb^D(C_{\boldsymbol{kl}}))\big) \\ &\le \sum_{\boldsymbol{k},\boldsymbol{l}} \big(\mu(C_{\boldsymbol{kl}}) - q_{\boldsymbol{kl}}\big) + \sum_{\boldsymbol{k},\boldsymbol{l}} \big|\mu(C_{\boldsymbol{kl}}) - h(0)Leb^D(C_{\boldsymbol{kl}})\big|, \end{aligned} \tag{5.13}$$

where the first sum was already estimated in (5.6). Its upper bound together with the upper bound from (5.12) forms the bound we arrived at for $d_2\big(\mathscr{L}(\mathrm{H}\tilde{\theta}_T^{-1}), \mathscr{L}(\eta\tilde{\theta}_T^{-1}|_{\tilde{J}_T})\big)$. Therefore, all that is left to do is to show that the second sum on the right hand side of Inequality (5.13) can be estimated by the claimed additional term. This, however, is

done very easily:

$$\begin{aligned}
\sum_{\boldsymbol{k},\boldsymbol{l}} \left|\mu(C_{\boldsymbol{kl}}) - h(0) Leb^D(C_{\boldsymbol{kl}})\right| &= \sum_{\boldsymbol{k},\boldsymbol{l}} \left| \int_{C_{\boldsymbol{kl}}} \big(h(s) - h(0)\big)\, \nu(ds\, dt) \right| \\
&\leq \int_{J_T} \left|h(s) - h(0)\right| \nu(ds\, dt) \\
&\leq 2^{D_2} L \cdot T \int_{\left[-(\frac{1}{w(T)})^{1/D_1}, (\frac{1}{w(T)})^{1/D_1}\right]^{D_1}} |s|^z \; Leb^{D_1}(ds) \\
&\leq 2^{D_2} D_1 L \alpha_{D_1} \cdot T \int_0^{\sqrt{D_1}(\frac{1}{w(T)})^{1/D_1}} r^{z+D_1-1}\, dr \\
&= 2^{D_2} D_1^{\frac{z+D_1}{2}} \frac{D_1}{z+D_1} L\alpha_{D_1} \frac{T}{w(T)^{1+z/D_1}}.
\end{aligned}$$

□

5.3 Applications

The results of the last section can be applied in a number of different ways. For one thing, they yield useful upper bounds for certain theoretical statements about Poisson process approximation, such as classical thinning and superposition theorems. This is explained in somewhat more detail in Subsection 5.3.1. But of course there are also more applied problems where the results of Section 5.2 can be of help. To obtain an idea of what is possible, we look at two examples from statistics in more detail: in Subsection 5.3.2 we consider a fairly general density estimation problem, examined also by Ellis (1991), and in Subsection 5.3.3 we consider a problem of testing for long range dependence.

5.3.1 Thinnings and superpositions

A brief introduction to thinnings and superpositions is given in Subsection 2.2.8. Since the problems of approximating these two point process modifications by a Poisson process are treated in detail in the next two chapters, we present here only the modeling framework without any explicit upper bounds. It can be seen that the results in Chapter 6 and 7 are in general clearly superior compared to what is possible here. Nevertheless, the following explanations point out an interesting link between the major approximation problems covered in this thesis.

Thinnings

Let $D_2 \geq 1$, and let ζ be a point process on $\mathbb{R}_+^{D_2}$ with locally finite expectation measure μ'. Suppose that $(p_n)_{n\in\mathbb{N}}$ is a sequence of measurable functions $\mathbb{R}_+^{D_2} \to [0,1]$ which is decreasing in n and that $\kappa_n : \mathbb{R}^{D_2} \to \mathbb{R}^{D_2}, s \mapsto \frac{1}{n}s$ denotes the contraction of space by the factor n. We are interested in the distributions of the thinned and contracted point processes $\zeta_{p_n}\kappa_n^{-1}|_{[-1,1]^{D_2}}$ for $n \in \mathbb{N}$.

In the notation of Sections 5.1 and 5.2, set $D_1 := 1$ and $w(T) := T$. Furthermore let P be a probability kernel from $(\mathbb{R}^{D_2}, \mathcal{B}^{D_2})$ to $(\mathbb{R}_+, \mathcal{B}_+)$ that satisfies $P(s, [0, 1/n^{D_2}]) = p_n(s)$

for all $s \in \mathbb{R}^{D_2}$ and $n \in \mathbb{N}$, where $\mathcal{B}^{D_2}$ and $\mathcal{B}_+$ denote the Borel σ-algebras on $\mathbb{R}^{D_2}$ and $\mathbb{R}_+$, respectively. Then, for a $\sigma(\zeta)$-measurable representation $\sum_{i=1}^V \delta_{S_i}$ of ζ, define the point process ξ on $\mathbb{R}^{D_1} \times \mathbb{R}^{D_2}$ to be $\xi := \sum_{i=1}^V \delta_{(Y_i,S_i)}$, where, given ζ, the Y_i are independent with distributions $P(S_i, \cdot)$. Thus ξ has a locally finite expectation measure μ, which can be calculated as

$$\mu(A) = \int P(s, A_s)\, \mu'(ds) = \big(\mu' \otimes P\big)(A)$$

for any $A \in \mathcal{B}_+ \otimes \mathcal{B}^{D_2}$, where $A_s := \{t \in \mathbb{R}_+;\, (t,s) \in A\}$ denotes the section of A at $s \in \mathbb{R}^{D_2}$.

It is easily seen that

$$\zeta_{p_n} \kappa_n^{-1}|_{[-1,1]^{D_2}} \stackrel{\mathscr{D}}{=} \big(\xi \tilde{\theta}_{T_n}^{-1}|_J\big)\mathrm{pr}_2^{-1}, \tag{5.14}$$

where $T_n := n^{D_2}$ and $\mathrm{pr}_2 : J = [-1,1]^{D_1} \times [-1,1]^{D_2} \to [-1,1]^{D_2}, (t,s) \mapsto s$. The following lemma is needed in order to dispose of the projection.

Lemma 5.3.A. *Let ξ_1 and ξ_2 be point processes on the compact space (E', d_0), and let $(\widetilde{E}', \tilde{d}_0)$ be another compact metric space whose metric is bounded by 1. Denote by $\tilde{d}_1$ and $\tilde{d}_2$ the relative Wasserstein metric and the Barbour-Brown metric with respect to $\tilde{d}_0$. Suppose furthermore that $\psi : E' \to \widetilde{E}'$ is Lipschitz continuous with constant 1. Then*

$$\tilde{d}_2\big(\mathscr{L}(\xi_1\psi^{-1}), \mathscr{L}(\xi_2\psi^{-1})\big) \le d_2\big(\mathscr{L}(\xi_1), \mathscr{L}(\xi_2)\big).$$

Proof. Let $\varrho_1, \varrho_2 \in \mathfrak{N}'$ be finite point measures on E'. If they do not have the same number of points, then obviously

$$\tilde{d}_1(\varrho_1\psi^{-1}, \varrho_2\psi^{-1}) = 1 = d_1(\varrho_1, \varrho_2).$$

If they do have the same positive number of points, choose representations $\varrho_1 = \sum_{i=1}^v \delta_{s_i}$ and $\varrho_2 = \sum_{i=1}^v \delta_{t_i}$ that are numbered in such a way that $\frac{1}{v}\sum_{i=1}^v d_0(s_i,t_i) = d_1(\varrho_1,\varrho_2)$. Because of the Lipschitz condition on ψ we have

$$\tilde{d}_1(\varrho_1\psi^{-1}, \varrho_2\psi^{-1}) \le \frac{1}{v}\sum_{i=1}^v \tilde{d}_0(\psi(s_i), \psi(t_i)) \le \frac{1}{v}\sum_{i=1}^v d_0(s_i,t_i) = d_1(\varrho_1,\varrho_2).$$

Hence, for the point processes ξ_1 and ξ_2,

$$\begin{aligned}
\tilde{d}_2\big(\mathscr{L}(\xi_1\psi^{-1}), \mathscr{L}(\xi_2\psi^{-1})\big) &= \min_{\substack{\xi_1' \sim \mathscr{L}(\xi_1) \\ \xi_2' \sim \mathscr{L}(\xi_2)}} \mathbb{E}\tilde{d}_1(\xi_1'\psi^{-1}, \xi_2'\psi^{-1}) \\
&\le \min_{\substack{\xi_1' \sim \mathscr{L}(\xi_1) \\ \xi_2' \sim \mathscr{L}(\xi_2)}} \mathbb{E} d_1(\xi_1', \xi_2') \\
&= d_2\big(\mathscr{L}(\xi_1), \mathscr{L}(\xi_2)\big).
\end{aligned}$$

□

We can now reduce the Poisson process approximation of an independent thinning to the situation of Sections 5.1 and 5.2. In order to give the Poisson processes involved explicitly, we compute the expectation measures for both sides of Equation (5.14). For

the left hand side we have, by using Campbell's formula (Proposition 2.2.E) for the last line,

$$\begin{aligned}\mathbb{E}\big(\zeta_{p_n}(B)\big) &= \mathbb{E}\bigg(\mathbb{E}\bigg(\sum_{i=1}^{V} X_i \delta_{S_i}(B) \;\bigg|\; \zeta\bigg)\bigg)\\ &= \mathbb{E}\bigg(\int_B p_n(s)\,\zeta(ds)\bigg)\\ &= \mathbb{E}\bigg(\int_B p_n(s)\,\mu'(ds)\bigg) =: \mu'_n(B),\end{aligned}$$

say, for any bounded $B \in \mathcal{B}$, where $(X_i)_{i\in\mathbb{N}}$ is a sequence of indicator random variables which are independent given ζ with $\mathbb{E}(X_i \,|\, \zeta) = p_n(S_i)$ for every i. Thus, the expectation measure of the left hand side in Equation (5.14) is $\mu'_n\kappa_n^{-1}|_{[-1,1]^{D_2}}$, whereas the expectation measure of the right hand side is obviously $\big(\mu\tilde{\theta}_{T_n}^{-1}|_J\big)\mathrm{pr}_2^{-1}$. Since these measures are equal (either by Equation (5.14) or by direct calculation), we have

$$\begin{aligned}d_2\Big(\mathscr{L}\big(\zeta_{p_n}\kappa_n^{-1}|_{[-1,1]^{D_2}}\big), \mathrm{Po}\big(\mu'_n\kappa_n^{-1}|_{[-1,1]^{D_2}}\big)\Big) &= d_2\Big(\mathscr{L}\big(\big(\xi\tilde{\theta}_{T_n}^{-1}|_J\big)\mathrm{pr}_2^{-1}\big), \mathrm{Po}\big(\big(\mu\tilde{\theta}_{T_n}^{-1}|_J\big)\mathrm{pr}_2^{-1}\big)\Big)\\ &\le d_2\Big(\mathscr{L}\big(\xi\tilde{\theta}_{T_n}^{-1}|_J\big), \mathrm{Po}\big(\mu\tilde{\theta}_{T_n}^{-1}|_J\big)\Big),\end{aligned} \tag{5.15}$$

where the inequality follows by Lemma 5.3.A, and d_2 denotes the Barbour-Brown metric with respect to the truncated Euclidean metric on J (and hence also on $[-1,1]^{D_2}$ by identification with $\{0\} \times [-1,1]^{D_2} \subset J$).

Thus it is possible to estimate the d_2 upper bound between a thinning law and a Poisson process law by a d_2 upper bound for distances of the form given in Theorem 5.2.A. The conditions needed for this theorem translate to similar but rather stronger conditions (as far as obtaining comparable upper bounds is concerned) than the ones we have in Chapter 6, where we consider Poisson process approximation of thinnings directly.

With the same idea it is also possible to model dependent thinnings, that is thinnings where the retention functions $\pi_n := p_n$ are random (see the part on thinnings in Subsection 2.2.8, or the formal thinning definition in Section 6.1). The probability kernel P from above would then be random, too (or following the alternative thinning approach of Section 6.4, new probability kernels would be introduced that model the dependence between the displacements Y_i of the points of ξ in the $\mathbb{R}^{D_1}$-directions directly). Moreover, in order to incorporate the thinning situation for all n in one model, we would essentially require the sequence (π_n) to be (pointwise) stochastically decreasing, meaning that for every $s \in \mathbb{R}^{D_2}$ and for $n \le n'$, we have $\pi_n(s) \overset{\mathscr{D}}{\ge} \pi_{n'}(s)$, that is the distribution function of $\pi_n(s)$ lies below the distribution function of $\pi_{n'}(s)$.

Superpositions

The superposition setting is quite a bit easier. However, we can deal directly only with the special case that there are finitely many point processes to superimpose, which become more and more sparse *by dilation of the state space*. Other sparseness mechanisms can be treated if the point process ξ from Theorem 5.2.A is allowed to depend on n, which in

principle is possible, because we have obtained explicit upper bounds for any fixed n in Inequalities (5.10) and (5.11).

Let $D_1 \geq 1$, and let $\zeta_1, \dots, \zeta_n$ be point processes on $\mathbb{R}^{D_1}$ with locally finite expectation measures $\mu_1, \dots, \mu_n$. Consider the dilations $\chi_n : \mathbb{R}^{D_1} \to \mathbb{R}^{D_1}, s \mapsto w(n)^{1/D_1} s$, for the $w(T)$ introduced in Section 5.1. We are interested in the distributions of the dilated superpositions $\big(\sum_{k=1}^n \zeta_k \chi_n^{-1}\big)\big|_{[-1,1]^{D_1}}$ for $n \in \mathbb{N}$.

In the notation of Sections 5.1 and 5.2, set $D_2 = 1$. Define the point process ξ on $\mathbb{R}^{D_1} \times \mathbb{R}^{D_2}$ to be $\xi := \sum_{k=1}^n \zeta_k \otimes \delta_k$, which has locally finite expectation measure $\mu = \sum_{k=1}^n \mu_k \otimes \delta_k$.

It is then easily seen that

$$\left(\sum_{k=1}^n \zeta_k \chi_n^{-1}\right)\Big|_{[-1,1]^{D_1}} = \big(\xi \tilde{\theta}_n^{-1}|_J\big)\mathrm{pr}_1^{-1}, \tag{5.16}$$

where $\mathrm{pr}_1 : J = [-1,1]^{D_1} \times [-1,1]^{D_2} \to [-1,1]^{D_1}, (s,t) \mapsto s$. Hence we can reduce Poisson process approximation of superpositions to the situation of Sections 5.1 and 5.2 in a similar way as in Inequality (5.15), by estimating

$$\begin{aligned} d_2\Big(\mathscr{L}\big(\big(\textstyle\sum_{k=1}^n \zeta_k \chi_n^{-1}\big)\big|_{[-1,1]^{D_1}}\big), \mathrm{Po}\big(\big(\textstyle\sum_{k=1}^n \mu_k \chi_n^{-1}\big)\big|_{[-1,1]^{D_1}}\big)\Big) \\ &= d_2\Big(\mathscr{L}\big(\big(\xi\tilde{\theta}_n^{-1}|_J\big)\mathrm{pr}_1^{-1}\big), \mathrm{Po}\big(\big(\mu\tilde{\theta}_n^{-1}|_J\big)\mathrm{pr}_1^{-1}\big)\Big) \\ &\leq d_2\Big(\mathscr{L}\big(\xi\tilde{\theta}_n^{-1}|_J\big), \mathrm{Po}\big(\mu\tilde{\theta}_n^{-1}|_J\big)\Big). \end{aligned}$$

Again, the conditions needed for Theorem 5.2.A translate to reasonable, but rather more restrictive results compared to the direct approximation of superpositions in Chapter 7. Note that the Poisson process used here for the approximation is not exactly the same as the one in Theorem 7.2.A, but is rather more natural because of the matching expectation measures. In order to arrive at this process in Chapter 7, we would estimate additionally the d_2-distance between two Poisson processes, in the same way as for the proof of Corollary 7.2.C.

5.3.2 Density estimation

First we give a new regularity condition for the density h of the expectation measure of ξ.

Assumption 4′ (Regularity of h). The density $h = d\mu/d\nu$ is constant in the $\mathbb{R}^{D_2}$-directions, so that we can write

$$h(s,t) = h(s) \quad \text{for all } s \in \mathbb{R}^{D_1}, t \in \mathbb{R}^{D_2} \text{ (resp. } t \in \mathbb{Z}^{D_2} + \tfrac{1}{2}\mathbf{1}).$$

Moreover, h satisfies the following regularity condition in the $\mathbb{R}^{D_1}$-directions:

$$h \in C^2(\mathbb{R}^{D_1}).$$

(In fact, it is enough if $h|_Z \in C^2(Z)$ for a sufficiently large neighborhood Z of $0 \in \mathbb{R}^{D_1}$.)

Suppose that Assumption 4′ holds, along with the usual conditions from Section 5.1. We would like to estimate the density h at the point $0 \in \mathbb{R}^{D_1}$, based on a single observation of ξ.

By way of illustration, it is convenient to think of the $\mathbb{R}^{D_1}$-space as the "data space" (i.e. the space of possible data points) and the $\mathbb{R}^{D_2}$-space as the "ascertainment space" (i.e. the space of points at which data is obtained, typically by continuous observation over time ($\mathbb{R}^{D_2} = \mathbb{R}$ = time axis) or by repetition of experiments ($\mathbb{R}^{D_2}$ with reference measure $\nu_2 = \mathcal{H}_0^{D_2}$)). An example suggested by Ellis (1986) and (1991) is the estimation of the rate at which earthquakes above a certain magnitude occur per unit area and unit time in a certain region. Here we have $D_1 = 2$ and $D_2 = 1$, and the points in $\mathbb{R}^3$ represent the positions and times of the observed earthquakes.

Among various methods for density estimation, we choose kernel estimation with a data-independent window width; that is, the window width in the $\mathbb{R}^{D_1}$-directions does not depend directly on the data, but does depend on the "observation span" (which in the discrete case corresponds to the sample size). For a detailed account of density estimation see Silverman (1986). We adapt the usual notation in connection with density estimation to the notation we used in Section 5.2. Thus, $2T^{1/D_2}$ is our observation span (in D_2 directions), $2/w(T)^{1/D_1}$ is the window width (in D_1 directions), and our density estimator at the point 0 takes the form

$$\hat{h}_\xi(0) := \frac{1}{|J_T|} \int_{J_T} 2^{D_1} K\big(w(T)^{1/D_1} s\big)\, \xi\big(d(s,t)\big),$$

where the function K is our kernel, which fulfills the following assumption.

Assumption 5 (Shape of K). The kernel $K : \mathbb{R}^{D_1} \to \mathbb{R}_+$ satisfies

(i) $K(s) = 0$ for $s \notin [-1,1]^{D_1}$;
(ii) $K|_{[-1,1]^{D_1}}$ is Lipschitz (w.r.t. d_0 restricted to $\mathbb{R}^{D_1}$) with constant $l(K)$;
(iii) $\int K(s)\, ds = 1$;
(iv) $\int K(s) s\, ds = 0 \in \mathbb{R}^{D_1}$.

Note that K does not have to be continuous on the boundary of $[-1,1]^{D_1}$, and that it is reasonable to choose a kernel K that is radially symmetric (or at least an even function in each coordinate), in which case Assumption 5(iv) is satisfied. We now write

$$K'(x) := 2^{D_1} K(s) \cdot 1_{[-1,1]^{D_2}}(t) \quad \text{for } x := (s,t) \in \mathbb{R}^{D_1} \times \mathbb{R}^{D_2} = \mathbb{R}^D,$$

so that $K'|_J$ is Lipschitz (w.r.t. d_0 on $\mathbb{R}^D$) with constant $2^{D_1} l(K)$; by the transformation theorem for integrals we obtain

$$\hat{h}_\xi(0) = \frac{1}{|J_T|} \int_{\mathbb{R}^D} K'(x)\, \xi\theta_T^{-1}(dx).$$

The way is now clear for the application of Theorem 5.2.A. Our primary goal will be to estimate a probability distance d between the distribution of our estimator $\hat{h}_\xi(0)$ and the distribution that is concentrated at the true value $h(0)$. To do this, we will first estimate $d\big(\mathscr{L}(\hat{h}_\xi(0)), \mathscr{L}(\hat{h}_\eta(0))\big)$ with the aid of Theorem 5.2.A, and then utilize the excellent properties of Poisson point processes to obtain an upper bound for $d\big(\mathscr{L}(\hat{h}_\eta(0)), \delta_{h(0)}\big)$.

The two corresponding results are contained in the following theorems. For the metric d we choose the bounded Wasserstein metric (which is the Wasserstein metric with respect to $|\cdot| \wedge 1$), because the other mertrics that we have used so far are too strong to be useful: $d_{TV}\big(\mathscr{L}(\hat{h}_\xi(0)), \delta_{h(0)}\big)$ is generally too big, and is even always equal to 1 whenever $\hat{h}_\xi(0)$ is a continuous random variable, because then

$$1 \geq d_{TV}\big(\mathscr{L}(\hat{h}_\xi(0)), \delta_{h(0)}\big) \geq \Big|\mathbb{P}[\hat{h}_\xi(0) = h(0)] - \mathbb{P}[h(0) = h(0)]\Big| = 1;$$

and for the Wasserstein distance $d_W\big(\mathscr{L}(\hat{h}_\xi(0)), \mathscr{L}(\hat{h}_\eta(0))\big)$ with respect to the Euclidean metric, an upper bound corresponding to that in Theorem 5.3.B is more involved and only useful under additional conditions.

Theorem 5.3.B. *Suppose that the prerequisites of Section 5.1 hold, including the Assumptions 1, 2, 3ϱ, as well as the additional Assumptions 4′ and 5 above. Let $\underline{c} > 0$, and for $T \geq 1$, let $m := m(T) \in \mathbb{Z}_+$, $\varsigma(T) \geq 1$ and also $w(T) = O(T^{\delta^*})$ for $T \to \infty$ with $\delta^* \in (0,1)$. Then*

$$\begin{aligned}
&d_{BW}\Big(\mathscr{L}\big(\hat{h}_\xi(0)\big), \mathscr{L}\big(\hat{h}_\eta(0)\big)\Big) \\
&\quad\leq \left(\frac{l(K)}{2^{D_2}}\frac{w(T)}{T}M + 1\right) d_2\Big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\eta\theta_T^{-1}|_J)\Big) + 2^{D_1} l(K)\delta_T(M) \\
&\quad= O\bigg(\frac{1}{\varsigma(T)^{1/D_1}}\,,\ \frac{1}{T^{1/D_2}}\,,\ \log^\uparrow\Big(\frac{T}{w(T)}\Big)\frac{m^{D_2}+1}{w(T)}\,, \\
&\qquad\quad \frac{T}{w(T)}\breve{\alpha}\Big(\frac{2^{D_2}}{w(T)\varsigma(T)}\Big)\,,\ \log^\uparrow\Big(\frac{T}{w(T)}\Big)\breve{\alpha}\Big(\frac{2^D(2m+1)^{D_2}}{w(T)}\Big)\,,\ \sqrt{T\varsigma(T)}\breve{\beta}(m)\bigg) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } T \to \infty,
\end{aligned}$$

where $M := M(T) \in \mathbb{N}$ with $M \geq 3\,\mu(J_T)$ arbitrary and

$$\delta_T(M) := \frac{\overline{c}}{\sqrt{M}}\Big(\frac{\mu(J_T)}{M}\Big)^M e^{M-\mu(J_T)},$$

which decays exponentially in M as T tends to infinity. Thus, we obtain the same order for the upper bound as in Theorem 5.2.A.

Remark 5.3.C. The upper bound given in Theorem 5.3.B remains true for general $w(T) = O(T)$. However, if $w(T)$ goes to infinity at a rate that is too close to T, then $M(T)$ has to be chosen to grow somewhat faster than $T/w(T)$, and then the order of the upper bound is a little worse (by a logarithmic factor in T) than the one stated in Theorem 5.3.B.

Proof. Let $\xi' \sim \mathscr{L}(\xi)$, $\eta' \sim \mathscr{L}(\eta) = \mathrm{Po}(\mu)$, and $X := \hat{h}_{\xi'}(0)$, $Y := \hat{h}_{\eta'}(0)$. Then we have

$$d_{BW}\Big(\mathscr{L}\big(\hat{h}_\xi(0)\big), \mathscr{L}\big(\hat{h}_\eta(0)\big)\Big) = \sup_{f \in \mathfrak{F}_{BW}} \big|\mathbb{E}f(X) - \mathbb{E}f(Y)\big|$$

with

$$\begin{aligned}\big|\mathbb{E}f(X)-\mathbb{E}f(Y)\big| &\le \mathbb{E}\Big(\big|f(X)-f(Y)\big|\,1_{\{\xi'\theta_T^{-1}(J)=\eta'\theta_T^{-1}(J)\}}\Big)\\ &\quad+\mathbb{E}\Big(\big|f(X)-f(Y)\big|\,1_{\{\xi'\theta_T^{-1}(J)\neq\eta'\theta_T^{-1}(J)\}}\Big)\\ &\le \mathbb{E}\Big(|X-Y|\,1_{\{\xi'\theta_T^{-1}(J)=\eta'\theta_T^{-1}(J)\}}\Big)\\ &\quad+\mathbb{P}\big[\xi'\theta_T^{-1}(J)\neq\eta'\theta_T^{-1}(J)\big] \qquad (5.17)\end{aligned}$$

for every f in $\mathfrak{F}_{BW}$. For the first summand we obtain

$$\begin{aligned}&\mathbb{E}\Big(|X-Y|\,1_{\{\xi'\theta_T^{-1}(J)=\eta'\theta_T^{-1}(J)\}}\Big)\\ &=\mathbb{E}\bigg(\frac{1}{|J_T|}\bigg|\int_{\mathbb{R}^D}K'(x)\;\xi'\theta_T^{-1}(dx)-\int_{\mathbb{R}^D}K'(x)\;\eta'\theta_T^{-1}(dx)\bigg|\,1_{\{\xi'\theta_T^{-1}(J)=\eta'\theta_T^{-1}(J)\}}\bigg)\\ &\le 2^{D_1}l(K)\,\mathbb{E}\bigg(\frac{\eta'\theta_T^{-1}(J)}{|J_T|}d_1\big(\xi'\theta_T^{-1}|_J,\eta'\theta_T^{-1}|_J\big)\bigg),\end{aligned}$$

the latter inequality by the definition of the d_1-distance and because $K'|_J$ is Lipschitz. Next we utilize the fact that since $\eta'\theta_T^{-1}(J)$ is Poisson distributed with parameter $\mu_T := \mu(J_T)$, it exceeds a certain bound $M := M(T) \in \mathbb{N}$ with $M \ge 3\mu_T$, say, only with very small probability. Lemma A.1.A in the appendix yields that the relation

$$\mathbb{P}\big[\eta'\theta_T^{-1}(J)\ge M\big]\le\frac{1}{\sqrt{2\pi M}}\,\frac{M+1}{M+1-\mu_T}\Big(\frac{\mu_T}{M}\Big)^M e^{M-\mu_T}\le\frac{1}{\sqrt{M}}\Big(\frac{\mu_T}{M}\Big)^M e^{M-\mu_T}$$

holds, and thus

$$\begin{aligned}&\mathbb{E}\bigg(\frac{\eta'\theta_T^{-1}(J)}{|J_T|}d_1\big(\xi'\theta_T^{-1}|_J,\eta'\theta_T^{-1}|_J\big)\bigg)\\ &\le\mathbb{E}\bigg(\frac{M}{|J_T|}d_1\big(\xi'\theta_T^{-1}|_J,\eta'\theta_T^{-1}|_J\big)1_{\{\eta'\theta_T^{-1}(J)\le M\}}\bigg)+\mathbb{E}\bigg(\frac{\eta'\theta_T^{-1}(J)}{|J_T|}1_{\{\eta'\theta_T^{-1}(J)>M\}}\bigg)\\ &\le\frac{M}{|J_T|}\mathbb{E}\Big(d_1\big(\xi'\theta_T^{-1}|_J,\eta'\theta_T^{-1}|_J\big)\Big)+\frac{\mu_T}{|J_T|}\mathbb{P}\big[\eta'\theta_T^{-1}(J)\ge M\big]\\ &\le\frac{1}{2^D}\frac{w(T)}{T}M\mathbb{E}\Big(d_1\big(\xi'\theta_T^{-1}|_J,\eta'\theta_T^{-1}|_J\big)\Big)+\delta_T(M)\end{aligned}$$

for $M \ge 3\mu_T$, where

$$\delta_T(M)=\frac{\overline{c}}{\sqrt{M}}\Big(\frac{\mu_T}{M}\Big)^M e^{M-\mu_T}\le\overline{c}\Big(\frac{e}{3}\Big)^M.$$

The second summand from Inequality (5.17) is estimated as

$$\begin{aligned}\mathbb{P}\big[\xi'\theta_T^{-1}(J)\neq\eta'\theta_T^{-1}(J)\big] &= \mathbb{E}\Big(d_1\big(\xi'\theta_T^{-1}|_J,\eta'\theta_T^{-1}|_J\big)1_{\{\xi'\theta_T^{-1}(J)\neq\eta'\theta_T^{-1}(J)\}}\Big)\\ &\le \mathbb{E}\Big(d_1\big(\xi'\theta_T^{-1}|_J,\eta'\theta_T^{-1}|_J\big)\Big).\end{aligned}$$

Hence we obtain altogether in (5.17)

$$\big|\mathbb{E}f(X) - \mathbb{E}f(Y)\big| \le \left(\frac{l(K)}{2^{D_2}}\frac{w(T)}{T}M + 1\right)\mathbb{E}\Big(d_1\big(\xi'\theta_T^{-1}|_J, \eta'\theta_T^{-1}|_J\big)\Big) + 2^{D_1}l(K)\delta_T(M)$$

for every $f \in \mathfrak{F}_{BW}$ and every pair of random variables ξ', η' with $\xi' \sim \mathscr{L}(\xi)$, $\eta' \sim \mathscr{L}(\eta)$. Forming the infimum over ξ' and η' yields on the right hand side the d_2-distance (θ_T is bijective), and forming the supremum over f on the left hand side the bounded Wasserstein distance. Thus we obtain the statement. □

The second result that was discussed above is contained in the next theorem. We write $\|\cdot\|_2$ for the L_2-norm with respect to the Lebesgue measure on $\mathbb{R}^{D_1}$.

Theorem 5.3.D. *Suppose that the prerequisites of Section 5.1 hold, including the Assumptions 1, 2, 3ϱ, as well as the additional Assumptions 4′ and 5 above. Let $\underline{c} > 0$, and for $T \ge 1$, let $m := m(T) \in \mathbb{Z}_+$, $\varsigma(T) \ge 1$ and also $w(T) = O(T^{\delta^*})$ for $T \to \infty$ with $\delta^* \in (0,1)$. Then*

$$\begin{aligned}
&d_{BW}\Big(\mathscr{L}\big(\hat{h}_\xi(0)\big), \delta_{h(0)}\Big) \\
&\le d_{BW}\Big(\mathscr{L}\big(\hat{h}_\xi(0)\big), \mathscr{L}\big(\hat{h}_\eta(0)\big)\Big) + \sqrt{\frac{\overline{c}}{2^{D_2}}}\|K\|_2\sqrt{\frac{w(T)}{T}} + \frac{L'}{w(T)^{2/D_1}} + o\Big(\frac{1}{w(T)^{2/D_1}}\Big) \\
&= O\bigg(\sqrt{\frac{w(T)}{T}},\ \frac{1}{w(T)^{2/D_1}},\ \frac{1}{\varsigma(T)^{1/D_1}},\ \frac{1}{T^{1/D_2}},\ \log^\uparrow\Big(\frac{T}{w(T)}\Big)\frac{m^{D_2}+1}{w(T)}, \\
&\qquad \frac{T}{w(T)}\breve{\alpha}\Big(\frac{2^{D_2}}{w(T)\varsigma(T)}\Big),\ \log^\uparrow\Big(\frac{T}{w(T)}\Big)\breve{\alpha}\Big(\frac{2^D(2m+1)^{D_2}}{w(T)}\Big),\ \sqrt{T\varsigma(T)}\breve{\beta}(m)\bigg) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } T \to \infty,
\end{aligned}$$

where L' is a non-negative constant (depending on h and K); if K possesses certain symmetry properties (especially if K is radially symmetric), we can write

$$L' := \frac{1}{2}\Delta h(0)\int s_1^2 K(s)\ Leb^{D_1}(ds),$$

where Δ denotes the D_1-dimensional Laplace operator.

Proof. Due to Theorem 5.3.B we only have to estimate $d_{BW}\big(\mathscr{L}(\hat{h}_\eta(0)), \delta_{h(0)}\big)$, where $\eta \sim \mathrm{Po}(\mu)$. We decompose this distance as

$$\begin{aligned}
d_{BW}\Big(\mathscr{L}\big(\hat{h}_\eta(0)\big), \delta_{h(0)}\Big) &\le d_{BW}\Big(\mathscr{L}\big(\hat{h}_\eta(0)\big), \delta_{\mathbb{E}\hat{h}_\eta(0)}\Big) + d_{BW}\big(\delta_{\mathbb{E}\hat{h}_\eta(0)}, \delta_{h(0)}\big) \\
&\le \mathbb{E}\big|\hat{h}_\eta(0) - \mathbb{E}\hat{h}_\eta(0)\big| + \big|\mathbb{E}\hat{h}_\eta(0) - h(0)\big| \\
&\le \mathrm{sd}\big(\hat{h}_\eta(0)\big) + \mathrm{bias}\big(\hat{h}_\eta(0)\big).
\end{aligned}$$

For the standard deviation we obtain

$$\begin{aligned}
\mathrm{sd}\big(\hat{h}_\eta(0)\big) &= \sqrt{\mathrm{var}\Big(\frac{1}{|J_T|}\int_{\mathbb{R}^D} K'(x)\ \eta\theta_T^{-1}(dx)\Big)} \\
&= \frac{1}{|J_T|}\sqrt{\int_{\mathbb{R}^D} K'^2(x)\ \mu\theta_T^{-1}(dx)} \\
&\le \frac{1}{|J_T|}\sqrt{\overline{c}_T\Big(\frac{1}{w(T)}\int_{\mathbb{R}^{D_1}} 2^{2D_1}K^2(s)\ Leb^{D_1}(ds)\Big)\,\nu_2\Big(\big[-T^{1/D_2},T^{1/D_2}\big]^{D_2}\Big)} \\
&\le \sqrt{\frac{\overline{c}}{2^{D_2}}}\|K\|_2\sqrt{\frac{w(T)}{T}},
\end{aligned}$$

where the second and third steps are applications of Corollary 2.2.P(ii) and Fubini's theorem, respectively (note that $(Leb^{D_1}\otimes\nu_2)\theta_T^{-1} = \frac{1}{w(T)}Leb^{D_1}\otimes\nu_2(T^{1/D_2}I_{D_2})$, where $I_{D_2}:\mathbb{R}^{D_2}\to\mathbb{R}^{D_2}$ is the identity). Furthermore, Corollary 2.2.P(i) and again Fubini's theorem yield

$$\begin{aligned}
&\mathbb{E}\hat{h}_\eta(0) \\
&\quad= \frac{1}{|J_T|}\int_{\mathbb{R}^D} K'(x)\ \mu\theta_T^{-1}(dx) \\
&\quad= \frac{1}{|J_T|}\Big(\frac{1}{w(T)}\int_{\mathbb{R}^{D_1}} 2^{D_1}K(s)h\Big(\frac{1}{w(T)^{1/D_1}}s\Big)\ Leb^{D_1}(ds)\Big)\,\nu_2\Big(\big[-T^{1/D_2},T^{1/D_2}\big]^{D_2}\Big) \\
&\quad= \int_{\mathbb{R}^{D_1}} K(s)h\Big(\frac{1}{w(T)^{1/D_1}}s\Big)\ Leb^{D_1}(ds).
\end{aligned}$$

Thus we obtain for the bias

$$\begin{aligned}
&\big|\mathbb{E}\hat{h}_\eta(0) - h(0)\big| \\
&\quad= \Big|\int_{[-1,1]^{D_1}} K(s)\Big(h\Big(\frac{1}{w(T)^{1/D_1}}s\Big) - h(0)\Big)\ Leb^{D_1}(ds)\Big| \\
&\quad\le \Big|\int_{[-1,1]^{D_1}} K(s)\frac{1}{w(T)^{1/D_1}}\partial h(0)s\ Leb^{D_1}(ds)\Big| \\
&\qquad+ \Big|\int_{[-1,1]^{D_1}} K(s)\frac{1}{2w(T)^{2/D_1}}\partial^2 h(0)(s,s)\ Leb^{D_1}(ds)\Big| \\
&\qquad+ \int_{[-1,1]^{D_1}} K(s)\frac{1}{2w(T)^{2/D_1}}\max_{0\le\zeta\le1}\Big\|\partial^2 h\Big(\zeta\frac{1}{w(T)^{1/D_1}}s\Big) - \partial^2 h(0)\Big\|\big|s\big|^2\ Leb^{D_1}(ds)
\end{aligned}$$

by Taylor approximation, where $\|\cdot\|$ is the standard norm for bilinear forms on $\mathbb{R}^{D_1}$. Of the last three summands, the first is always zero because of Assumption 5(iv), the second can be estimated by $L'\frac{1}{w(T)^{2/D_1}}$ with a constant L', which for "nice" kernels (e.g. if K is radially symmetric) can be written as

$$L' = \frac{1}{2}\Delta h(0)\int s_1^2 K(s)\ Leb^{D_1}(ds),$$

and the third is of order $o\big(\frac{1}{w(T)^{2/D_1}}\big)$ because of the continuity of $\partial^2 h$ at 0. Thus

$$\mathrm{bias}\big(\hat{h}_\eta(0)\big) \le L' \frac{1}{w(T)^{2/D_1}} + o\Big(\frac{1}{w(T)^{2/D_1}}\Big).$$

□

Once more we formulate the conditions under which the upper bound goes to zero.

Corollary 5.3.E (Convergence to zero in Theorem 5.3.D). *Suppose that the prerequisites of Theorem 5.3.D hold. Furthermore, suppose that* $w(T) \ge kT^\delta$ *for* $k > 0$, $\delta \in (0,1)$ *and that*

$$\breve{\alpha}(v) = O\big(v^r\big) \quad \text{for } v \to 0 \qquad \text{with } r > 0,$$
$$\breve{\beta}(u) = O\Big(\frac{1}{u^{(1+s)D_2/2}}\Big) \quad \text{for } u \to \infty \qquad \text{with } 1 + s > \max\Big(\frac{1-\delta}{\delta}\frac{1+r}{r},\ \frac{1}{\delta}\Big).$$

Then

$$d_{BW}\Big(\mathscr{L}\big(\hat{h}_\xi(0)\big), \delta_{h(0)}\Big) \longrightarrow 0 \quad \text{for } T \to \infty,$$

and therefore, since by Theorem 2.4.D the d_{BW}*-distance metrizes convergence in distribution and since* $\delta_{h(0)}$ *is the distribution of a constant, we obtain*

$$\hat{h}_\xi(0) \xrightarrow{P} h(0) \quad \text{for } T \to \infty,$$

that is, the consistency of the estimator $\hat{h}_\xi(0)$.

Remark 5.3.F. The consistency of $\hat{h}_\xi(0)$ was already obtained as a consequence of Theorem 2.5 in Ellis (1991) under conditions that were similar, but for the most part somewhat more general. So Corollary 5.3.E is not so much a new result, but rather a crosscheck on the suitability of the explicit upper bound obtained in Theorem 5.3.D.

Proof. Let $M := \lceil 3\,\mu(J_T) \rceil$ in Theorem 5.3.B. We then get immediately by applying Theorem 5.3.D, Theorem 5.3.B and Corollary 5.2.B that $d_{BW}\big(\mathscr{L}(\hat{h}_\xi(0)), \delta_{h(0)}\big)$ converges to zero. □

5.3.3 Testing for long range dependence

Suppose that ξ is a stationary point process on $\mathbb{R}^D$ with expectation measure $\mu = \mu_0 Leb^D$ ($\mu_0 > 0$ known or estimated) which satisfies the conditions of Section 5.1, except for Assumption 3. We would like to test from a single realization of ξ whether there is important long range dependence in the $\mathbb{R}^{D_2}$-directions or not (our null hypothesis). "No important long range dependence" means here that Assumption $3x$ is satisfied for given $x \in \{\beta, \varrho, \varphi\}$ and $\breve{\beta}$, corresponding to the minimal mixing rate one wants to test for. For the sake of illustration, think of the $\mathbb{R}^{D_1}$-direction(s) as time, and the $\mathbb{R}^{D_2}$-directions as space. Imagine that, for fixed $T \ge 1$, the points of ξ in J_T denote the times and locations of occurrences of cases of a certain rare disease, which is observed in a large area (e.g. a country or a continent) over a relatively short period of time (e.g. some months or a year).

Under the null hypothesis, by Theorem 5.2.A resp. Theorem 5.2.D, the distribution of $\xi\theta_T^{-1}|_J$ will be close to the distribution of $\eta\theta_T^{-1}|_J$, which here is just the homogeneous Poisson process on J with intensity $\mu_0\frac{T}{w(T)}$. There are various reasonable statistics for testing the hypothesis of "complete spatial randomness" in point patterns. One such statistic $U : \mathfrak{N}' \to \mathbb{R}_+$ is the average nearest neighbor distance in the data, which has been introduced as Example 7 in Subsection 3.3.1. It was shown there that U is Lipschitz continuous with constant $\tau_D + 1$, where τ_D denotes the maximal kissing number in D dimensions.

We wish to find an approximate critical value t_α for, say, a one-sided test of size α of the null hypothesis against an aggregated alternative (i.e. the alternative that there is a certain amount of "long range" clustering), using the statistic $\tilde{U}$, where $\tilde{U}(\varrho) := U(\varrho\theta_T^{-1}|_J)$ for every point measure ϱ on $\mathbb{R}^D$. To do so, fix $M > 0$ and choose t_α so that

$$\mathbb{E}f_{t_\alpha,M}\big(\tilde{U}(\eta)\big) + M(\tau_D + 1)\varepsilon = \alpha,$$

where ε is our upper bound for $d_2\big(\mathscr{L}(\xi\theta_T^{-1}|_J), \mathscr{L}(\eta\theta_T^{-1}|_J)\big)$, and

$$f_{t,M}(x) := \begin{cases} 1 & \text{if } x \le t \\ 1 - M(x-t) & \text{if } t \le x \le t + \frac{1}{M} \\ 0 & \text{if } x \ge t + \frac{1}{M} \end{cases}$$

is an M-Lipschitz approximation of the indicator $1_{(-\infty,t]}$. This yields

$$0 \;\le\; \alpha - \mathbb{P}\big[\tilde{U}(\xi) < t_\alpha\big] \;\le\; \mathbb{E}f_{t_\alpha,M}\big(\tilde{U}(\eta)\big) - \mathbb{E}f_{(t_\alpha - 1/M),M}\big(\tilde{U}(\eta)\big) + 2M(\tau_D + 1)\varepsilon.$$

Thus, if ε is very small (i.e. the conditions for Theorem 5.2.A resp. Theorem 5.2.D are strong enough), a large M can be chosen, and consequently we can adjust the size of our test to be only slightly below α.

It should be noted that the distribution of $\tilde{U}(\eta)$ is not known, but it can be simulated very easily. Also, there are good normal approximations to $\mathscr{L}\big(\tilde{U}(\eta) \;\big|\; |\eta| = v\big)$ for v not too small which can be of use. See Ripley (1981), Section 8.2, for further details.

Chapter 6

Approximating thinnings of point processes

It is well known that, under certain conditions, gradual thinning of a point process on $E := \mathbb{R}_+^D$, accompanied by a contraction of space to compensate for the thinning, leads in the weak limit to a Cox process. In this chapter we apply methods similar to those in the last one, especially discretization and the local Stein Theorem 4.3.A, in order to give estimates of the d_2-distance between the distribution of a thinned point process and an approximating Poisson process, and we evaluate the estimates in concrete examples. We work in terms of two somewhat different thinning models. The main model is based on the usual thinning notion of deleting points independently according to probabilities supplied by a random field. This model was briefly introduced in Subsection 2.2.8. However, we also use an alternative thinning model, which can be more straightforward to apply if the thinning is determined by point interactions (see Section 6.4).

We start by giving the formal thinning definition that was announced in Subsection 2.2.8, along with some general notation and conventions.

6.1 Thinning definition and some conventions

In all of Chapter 6 we assume that the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is complete in order to avoid measurability problems. Denote the Borel σ-field on $E = \mathbb{R}_+^D$ by $\mathcal{B}_+^D$ and the trace σ-field $\mathcal{B}_+^D|_A$ as usual by $\mathcal{B}_A$ for any set $A \subset \mathbb{R}_+^D$.

In view of the Barbour-Brown metric we restrict our point processes this time to the unit cube $E' := J := [0,1]^D$. Furthermore write $J_n := [0,n]^D$, $J_n' := [0,n)^D$, and $C_{\boldsymbol{k}} := \prod_{i=1}^D [k_i - 1, k_i)$ for all $\boldsymbol{k} = (k_1, \ldots, k_D) \in \{1, 2, \ldots, n\}^D$ for the unit hypercubes that make up J_n'. As in Chapter 5, we make use of a simplified multi-index notation by denoting properties of all the individual components as though they were properties of the whole multi-index, for example writing $J_n' = \bigcup_{\boldsymbol{k}=1}^n C_{\boldsymbol{k}}$ instead of $J_n' = \bigcup_{\boldsymbol{k}:k_1,\ldots k_D=1}^n C_{\boldsymbol{k}}$. Furthermore, we use once more the distance between multi-indices that is defined by $|\boldsymbol{l} - \boldsymbol{k}| := \max_{1 \le i \le D} |l_i - k_i|$. Recall also the convention from Section 2.1, that the addition "almost surely" may be left away in equations and inequalities between random elements in situations where it is evident (and unimportant) that the corresponding relation does not hold pointwise.

In what follows, ξ is always a point process on $\mathbb{R}_+^D$ and $\pi := \{\pi(\cdot, s); s \in \mathbb{R}_+^D\}$ a $[0,1]$-valued random field on $\mathbb{R}_+^D$ that satisfies the following measurability condition: for any bounded rectangle R that is open in $\mathbb{R}_+^D$, the mapping $\pi_R^* : \Omega \times R \to [0,1]$, $(\omega, s) \mapsto \pi(\omega, s)$ is $\sigma(\pi|_R) \otimes \mathcal{B}_R$-measurable (we say in this case that π *is locally evaluable*). Assuming this technical condition considerably simplifies some of the notation and the arguments in Section 6.3. However, if one accepts a more involved presentation, in particular larger and more complicated σ-fields in Assumption 2 below, and a stronger independence property for Theorem 6.3.F, all that is needed to prove the corresponding theorems is the measurability of π as a mapping $\Omega \times \mathbb{R}_+^D \to [0,1]$. The latter is necessary to ensure that for any random point S in $\mathbb{R}_+^D$, $\pi(S)$ is a random variable. Note that local evaluability is satisfied for many desirable random fields, as for example, for random fields which are pathwise continuous or for indicators of (fully) separable random closed sets (see Appendix A.3 for a more detailed discussion of locally evaluable random fields, with proofs).

The π-thinning of ξ can now be defined as follows.

Definition (Thinning). First, assume that $\xi = \sigma = \sum_{i=1}^{v} \delta_{s_i}$ and $\pi = p$ are non-random, where $v \in \overline{\mathbb{Z}}_+$, $s_i \in \mathbb{R}_+^D$, and p is a function $\mathbb{R}_+^D \to [0,1]$. Then, a π-thinning of ξ is defined as $\xi_\pi = \sum_{i=1}^{v} X_i \delta_{s_i}$, where the X_i are independent indicator random variables with expectations $p(s_i)$, respectively. Under these circumstances, ξ_π has a distribution $P_{(\sigma,p)}$ that does not depend on the chosen enumeration of σ. For general ξ and π, define then a π-*thinning* ξ_π by randomization, that is by the condition $\mathbb{P}[\xi_\pi \in \cdot \,|\, \xi, \pi] = P_{(\xi,\pi)}$ (under our conditions on π it is straightforward to see that $P_{(\xi,\pi)}$ is a $\sigma(\xi,\pi)$-measurable family in the sense that $P_{(\xi,\pi)}(N)$ is $\sigma(\xi,\pi)$-measurable for every $N \in \mathcal{N}$). Note that the distribution of ξ_π is uniquely determined by this procedure.

This thinning definition can be found in Kallenberg (1986) for non-random π, and is generalized in Serfozo (1984) to random π. The definition of Brown (1979), and the less formal definitions of Stoyan, Kendall and Mecke (1987) and Daley and Vere-Jones (1988) also yield the same distribution for ξ_π. Recall from Subsection 2.2.8 that we usually refer to ξ as "the original process" and to π as "the retention field" in the thinning context.

The following remark simplifies the presentation of the proofs.

Remark 6.1.A (Definition of retention decisions). Given a countable partition $(B_j)_{j\in\mathbb{N}}$ of $\mathbb{R}_+^D$ into bounded measurable sets, there is by Corollary 2.2.D always a representation of ξ as $\sum_{i=1}^{\xi(\mathbb{R}_+^D)} \delta_{S_i}$ with $\mathbb{R}_+^D$-valued, $\sigma(\xi)$-measurable random elements S_i such that $S_1, \ldots, S_{\xi(B_1)} \in B_1$, $S_{\xi(B_1)+1}, \ldots, S_{\xi(B_1)+\xi(B_2)} \in B_2$, and so on. We will make tacit use of this fact in connection with the thinning definition on various occasions. For example, for a point process ξ and a bounded Borel set A, we may write $\xi_\pi(A) = \sum_{i=1}^{\xi(A)} X_i$, and hereby imply that we define $\sigma(\xi)$-measurable "point random elements" S_i for ξ in such a way that the first $\xi(A)$ points $S_1, \ldots, S_{\xi(A)}$ always lie in A, and all other points in A^c, and define "retention decisions" X_i which, conditional on ξ and π_n, are independent with expectations $\pi_n(S_i)$, respectively.

6.2 Introduction to the approximation problem

Consider a sequence $(\pi_n)_{n\in\mathbb{N}}$ of locally evaluable retention fields with $\sup_{s\in\mathbb{R}_+^D} \pi_n(s) \xrightarrow{\mathscr{D}} 0$, corresponding to the idea of ξ being gradually thinned away. To compensate for this effect,

we contract the Euclidean space by $\kappa_n : \mathbb{R}_+^D \to \mathbb{R}_+^D$, $s \mapsto \frac{1}{n}s$. Then, the following is a standard result.

Theorem 6.2.A. *We obtain convergence in distribution of the thinned and contracted process $\xi_\pi \kappa_n^{-1}$ to a point process η if and only if, with $\Lambda_n(A) := \int_{\kappa_n^{-1}(A)} \pi_n(s)\, \xi(ds)$ for every measurable set $A \subset \mathbb{R}_+^D$, we have*

$$\Lambda_n \xrightarrow{\mathscr{D}} \Lambda \quad \text{for } n \to \infty \tag{6.1}$$

for some random measure Λ on $\mathbb{R}_+^D$. In this case $\eta \sim \mathrm{Cox}(\Lambda)$; that is, η is a Cox process with directing measure Λ.

Proof. See Daley and Vere-Jones (1988), Theorem 9.3.III (of course, measurability of π_n is enough). The proof is done with the aid of generating functionals. □

The main goal of this chapter is to examine the rate of convergence in the above theorem in the Barbour-Brown metric. Due to the method of proof that we use, we will always assume that a mixing condition holds for the random measures Λ_n (see Assumption 2 in Section 6.3). Since Condition (6.1) can be interpreted as the statement of a weak ergodic theorem, it is natural, in view of the usual chain "*mixing* implies *ergodic* implies *constant limit in the ergodic theorem*", that we get a deterministic limiting measure $\Lambda = \lambda$, and hence even a Poisson process as the limiting process η of the contracted thinning. These heuristics can easily be made rigorous if ξ is stationary, and $\pi_n = (1/n^D)\tilde{\pi}$ for a stationary random field $\tilde{\pi}$.

The thinning of point processes was first studied in Rényi (1957): a renewal process on $\mathbb{R}_+$ was subjected to an independent thinning with constant retention probability p, and a Poisson limit theorem was obtained for $p \to 0$ upon a change of time scale by a factor $1/p$. There have been many generalizations within the theory of point processes on the real line since then, with some of the most comprehensive found in Jagers and Lindvall (1974), Serfozo (1977), and Böker and Serfozo (1983).

Also in Kallenberg (1975) (alternatively, see Kallenberg (1986), Section 8.3), independent thinnings with constant retention probability p were considered, but this time the processes to be thinned were arbitrary point processes on general locally compact, second countable, Hausdorff spaces (lcscH). Necessary and sufficent conditions were derived for the convergence of thinnings of increasingly "dense" point processes to a Cox limit. This result was generalized in Brown (1979) to position dependent, random retention probabilities, which yielded, up to some negligible details in the setting, exactly the statement of Theorem 9.3.III in Daley and Vere-Jones (1988), from which Theorem 6.2.A is a direct consequence.

A result regarding distance estimates in thinning theorems may also be found in Daley and Vere-Jones (1988). In Proposition 9.3.IV, the authors give a quite abstract upper bound for the total variation distance between the distributions of the point counts $\xi_{\pi_n}\kappa_n^{-1}(A)$ and $\eta(A)$ for any bounded Borel set A. By a rather similar argument, it is possible to obtain a corresponding upper bound for the d_2-distance between the distributions of the restricted point processes $\xi_{\pi_n}\kappa_n^{-1}|_C$ and $\eta|_C$, where $C \subset \mathbb{R}_+^D$ is an arbitrary compact set. The proof of the following result can be found in Appendix A.4.1.

Proposition 6.2.B. *In the notation of Theorem 6.2.A we have for compact $C \subset \mathbb{R}_+^D$*

$$d_2\Big(\mathscr{L}\big(\xi_{\pi_n}\kappa_n^{-1}|_C\big), \mathrm{Cox}\big(\Lambda|_C\big)\Big) \leq \mathbb{E}\Bigg(6.5 \cdot \sup_{s\in\kappa_n^{-1}(C)} \pi_n(s) + \min\big(1, \Lambda_n(C), \Lambda(C)\big)\, d_W\Big(\frac{\Lambda_n|_C(\cdot)}{\Lambda_n(C)}, \frac{\Lambda|_C(\cdot)}{\Lambda(C)}\Big) + 1 \wedge \Bigg[\min\Bigg(1, \frac{1.65}{\sqrt{\Lambda_n(C)}}, \frac{1.65}{\sqrt{\Lambda(C)}}\Bigg)\Big|\Lambda_n(C) - \Lambda(C)\Big|\Bigg]\Bigg), \quad (6.2)$$

where the second summand is defined to be zero if either $\Lambda_n(C)$ or $\Lambda(C)$ is zero. Here, d_W denotes the Wasserstein metric with respect to d_0 on $\mathbb{R}_+^D$, which is the same as the bounded Wasserstein metric with respect to the Euclidean metric on $\mathbb{R}_+^D$. □

While such an upper bound is very nice from a theoretical point of view, because it is of a comparatively simple form, and because it can be shown that it goes to zero under the general Condition (6.1) if C satisfies $\mathbb{P}[\Lambda(\partial C) > 0] = 0$, it is usually not so easy to calculate. In what follows, we are interested in more explicit upper bounds for the thinning approximation, i.e. upper bounds that can be calculated directly in terms of certain characteristic quantities of the point process ξ and the retention field π. Since, in the (still quite general) case where ξ and π satisfy mixing conditions, and the approximating process is Poisson, we have available well-developed tools from the field of Stein's method for point process approximation, we prefer to make use of these, rather than trying to reduce the right hand side of Inequality (6.2) to more fundamental characteristics of ξ and π.

Section 6.3 contains the main results, namely upper bounds for the d_2-distance between $\mathscr{L}(\xi_{\pi_n}\kappa_n^{-1}|_J)$ and a Poisson process law under different conditions. In Section 6.4, a slightly different notion of a thinning, called Q-thinning, is introduced and a corresponding upper bound is given.

The difference between the thinning model used up to Section 6.3 and the model used in Section 6.4 is conceptual rather than a difference in terms of the modeled objects (see Remarks 6.4.B and 6.4.E for details on how the resulting thinnings differ). Under π-thinning, points are more or less likely to be deleted according to the conditions they encounter in a random environment (which itself may respond to the point configuration); under Q-thinning, points are more or less likely to be deleted according directly to the interactions among themselves. This difference is nicely illustrated by the two applications given in Section 6.5. In both situations, we start with a point process ξ on $\mathbb{R}_+^D$ having "reasonable" first and second factorial moment measures, and obeying an appropriate mixing condition. We have a "basic" thinning effect given by the constant retention probability $q_0^{(n)}$, and an additional "characteristic" thinning effect for each of the two examples. The basic and the characteristic effect are combined, and $\mathbb{R}_+^D$ is contracted by a factor $1/n$ to obtain a point process that is compared to a Poisson process.

In the first example (Subsection 6.5.1), we consider a random environment given by a union Ξ of balls whose centers form a stationary Poisson process on $\mathbb{R}^D$, and whose radii are i.i.d. and bounded with L_D-norm r_n. The characteristic thinning effect is then provided by deleting all the points that are covered by Ξ (we illustrate this situation with the image of stars that are covered by clouds). In the second example (Subsection 6.5.2), we give

point interactions by assigning i.i.d. real-valued marks to the points. The characteristic thinning effect is then provided by deleting all the points whose distance to some other point having a larger mark is at most r_n (we illustrate this situation with the scenario of competition within a plant population).

Roughly speaking, Poisson approximation is good in both examples if the retention probabilities $q_0^{(n)}$ become appropriately small, or the ranges r_n become appropriately large.

It should be noted that the illustrations of the examples as "visibility of stars" and "plant competition", respectively, may well provide inspiration for modeling similar situations, but are not meant to be serious modeling attempts in themselves.

We end this section by giving an indication of the type of results that are obtained in Section 6.3. The following proposition covers the important special case when $\pi_n = p_n$ is non-random, which follows directly from either Theorem 6.3.C or Theorem 6.3.F. This type of thinning was introduced in Subsection 2.2.8 as independent thinning, because, given ξ, the retention decisions for the various points are independent of one another.

Proposition 6.2.C (Independent thinning). *Let ξ be a point process on $\mathbb{R}_+^D$ which has an expectation measure μ_1 that has a bounded density with respect to Lebesgue measure, and a second factorial moment measure $\mu_{[2]}$ that is bounded on the set of all unit cubes in $\mathbb{R}_+^{2D}$. Let $(p_n)_{n\in\mathbb{N}}$ be a sequence of functions $\mathbb{R}_+^D \to [0,1]$, and let $\bar{p}_n := \sup_{s\in\mathbb{R}_+^D} p_n(s)$. Suppose that "long range" covariances in ξ are controlled by a decreasing function $\breve{\beta} : \mathbb{R}_+ \to \mathbb{R}_+$, such that for every open cube $A_{\text{int}} = (a, a+h\mathbf{1})$ of side length $h \in (0,1]$ and with minimal corner at $a \in \mathbb{R}_+^D$, and every surrounding set $A_{\text{ext}}^{(t)} := \mathbb{R}_+^D \setminus [a - t\mathbf{1}, a + (t+h)\mathbf{1}]$ with $t \in \mathbb{R}_+$, we have*

$$\sup_{F,Z} \bigl|\operatorname{cov}(Z, 1_F)\bigr| \le \breve{\beta}(t),$$

where the supremum ranges over all sets $F \in \sigma\bigl(\xi|_{A_{\text{ext}}^{(t)}}\bigr)$ and all random variables $Z \in L_2\bigl(\sigma(\xi|_{A_{\text{int}}})\bigr)$ with $0 \le Z \le \xi(A_{\text{int}})/|A_{\text{int}}|$. Then, for arbitrary $m := m(n) \in \mathbb{N}$, we obtain

$$d_2\Bigl(\mathscr{L}\bigl(\xi_{p_n}\kappa_n^{-1}|_J\bigr), \operatorname{Po}\bigl(\lambda_n\kappa_n^{-1}|_J\bigr)\Bigr) = O\Bigl(n^D\bar{p}_n\bigl(m^D\bar{p}_n \vee \breve{\beta}(m)\bigr)\Bigr) \quad \text{for } n \to \infty$$

with $\lambda_n(\cdot) := \int_\cdot p_n(s)\,\mu_1(ds)$. In the most natural case, where $\bar{p}_n = O(1/n^D)$, this implies that

$$d_2\Bigl(\mathscr{L}\bigl(\xi_{p_n}\kappa_n^{-1}|_J\bigr), \operatorname{Po}\bigl(\lambda_n\kappa_n^{-1}|_J\bigr)\Bigr) = O\bigl(n^{-sD/(s+1)}\bigr)$$

if $\breve{\beta}(t) = O(t^{-sD})$ as $t \to \infty$ for some fixed $s > 0$.

Note that we maintain a great deal of flexibility, and avoid evaluating awkward distances between general probability measures, by approximating the thinning with a Poisson process whose intensity measure depends on n. If we must have a fixed approximating Poisson process, we get by Theorem 4.3.D the additional term of

$$\begin{aligned} d_2\Bigl(\operatorname{Po}\bigl(\lambda_n\kappa_n^{-1}|_J\bigr), \operatorname{Po}(\lambda|_J)\Bigr) \le{}& \min\bigl(1, \lambda_n\kappa_n^{-1}(J), \lambda(J)\bigr)\, d_W\Bigl(\frac{\lambda_n\kappa_n^{-1}|_J(\cdot)}{\lambda_n\kappa_n^{-1}(J)}, \frac{\lambda|_J(\cdot)}{\lambda(J)}\Bigr) \\ &+ \min\Bigl(1, \frac{1.65}{\sqrt{\lambda_n\kappa_n^{-1}(J)}}, \frac{1.65}{\sqrt{\lambda(J)}}\Bigr)\Bigl|\lambda_n\kappa_n^{-1}(J) - \lambda(J)\Bigr|, \quad (6.3) \end{aligned}$$

which is, in most cases, still much more convenient to estimate than the two corresponding terms in (6.2), inasmuch as the quantities appearing in (6.3) are no longer random.

6.3 The main results

For this whole section, let ξ be a point process on $\mathbb{R}_+^D$, and, for each $n \in \mathbb{N}$, let $\pi_n := \{\pi_n(s)\,;\, s \in \mathbb{R}_+^D\} := \{\pi_n(\cdot, s)\,;\, s \in \mathbb{R}_+^D\}$ be a $[0,1]$-valued, locally evaluable random field on $\mathbb{R}_+^D$. The following two assumptions, in one or the other form, are used several times in this chapter.

Assumption 1 (Control of moment measures).

a) ξ has an expectation measure $\mu_1 \ll Leb^D$ with bounded density $h_1 : \mathbb{R}_+^D \to \mathbb{R}_+$, and $\bar{h}_1 := \|h_1\|_\infty$;

b) ξ has a second factorial moment measure $\mu_{[2]}$ which is bounded on the set of unit cubes in $\mathbb{R}_+^{2D}$, i.e. there is $\bar{h}_2 \geq 0$ such that $\mu_{[2]}(C) \leq \bar{h}_2$ for any unit cube $C \subset \mathbb{R}_+^{2D}$.

Assumption 2 (Mixing property). For each $n \in \mathbb{N}$, let $\breve{\beta}_n : \mathbb{R}_+ \to \mathbb{R}_+$ be a decreasing function such that for every cube $A_{\text{int}} = (a, a + h\mathbf{1})$ of side length $h \in (0,1]$ and with minimal corner at $a \in \mathbb{R}_+^D$, and every surrounding set $A_{\text{ext}}^{(t)} := \mathbb{R}_+^D \setminus [a - t\mathbf{1}, a + (t+h)\mathbf{1}]$ with $t \in \mathbb{R}_+$, we have that

$$\sup_{\substack{F \in \mathcal{F}_{\text{ext}}^{(n,t)} \\ Z \in L_2(\mathcal{F}_{\text{int}}^{(n)}),\, 0 \leq Z \leq \Pi}} \big|\operatorname{cov}\big(Z, 1_F\big)\big| \leq \breve{\beta}_n(t),$$

where

$$\mathcal{F}_{\text{int}}^{(n)} := \sigma\big(\xi|_{A_{\text{int}}}, \pi_n|_{A_{\text{int}}}\big), \quad \mathcal{F}_{\text{ext}}^{(n,t)} := \sigma\big(\xi|_{A_{\text{ext}}^{(t)}}, \pi_n|_{A_{\text{ext}}^{(t)}}\big),$$

and

$$\Pi := \Big(\xi(A_{\text{int}}) \sup_{s \in A_{\text{int}}} \pi_n(s)\Big) \Big/ \mathbb{E}\Big(\xi(A_{\text{int}}) \sup_{s \in A_{\text{int}}} \pi_n(s)\Big)$$

if $\mathbb{E}\big(\xi(A_{\text{int}}) \sup_{s \in A_{\text{int}}} \pi_n(s)\big) > 0$, and $\Pi :\equiv 0$ otherwise.

Remark 6.3.A. Since π_n is a measurable random field and $\mathcal{A}$ is complete, it can be shown by a standard argument involving analytic sets (using the Lusin-Choquet-Meyer theorem from Kallenberg (2002), Appendix A1) that $\sup_{s \in A} \pi_n(s)$ is a random variable for any $A \in \mathcal{B}_+^D$.

Results

We start with the main theorem, first in its most general form, and then in weaker but less involved versions. Remember that we use $O\big(f_1(n), \ldots, f_j(n)\big)$ as short hand for $O\big(\max\{f_1(n), \ldots, f_j(n)\}\big)$. Quantitative versions of the upper bounds can be found in the proofs starting after Theorem 6.3.F.

Theorem 6.3.B (Principal thinning theorem). *Suppose that the point process ξ and the sequence $(\pi_n)_{n \in \mathbb{N}}$ of retention fields satisfy Assumptions 1 and 2 above. Set*

$$w_1 := \sup_{1 \leq \boldsymbol{k} \leq n} \mathbb{E}\Big(\xi(C_{\boldsymbol{k}}) \sup_{s \in C_{\boldsymbol{k}}} \pi_n(s)\Big),$$

$$w_{[2]} := \sup_{1 \leq \boldsymbol{k}, \boldsymbol{l} \leq n} \mathbb{E}\Big(\xi(C_{\boldsymbol{k}})\big(\xi(C_{\boldsymbol{l}}) - \delta_{\boldsymbol{kl}}\big) \sup_{s \in C_{\boldsymbol{k}}} \pi_n(s) \sup_{s \in C_{\boldsymbol{l}}} \pi_n(s)\Big),$$

$$\underline{w}_1 := \inf_{1\le \boldsymbol{k}\le n} \mathbb{E}\Big(1_{\{\xi(C_{\boldsymbol{k}})\ge 1\}} \inf_{s\in C_{\boldsymbol{k}}} \pi_n(s)\Big),$$

and let $\lambda_n(\cdot) := \mathbb{E}\big(\int_\cdot \pi_n(s)\ \xi(ds)\big)$*, the expectation measure of* ξ_{π_n}*. Then, for arbitrary* $m := m(n) \in \mathbb{N}$*, we obtain*

$$d_2\Big(\mathscr{L}\big(\xi_{\pi_n}\kappa_n^{-1}|_J\big), \mathrm{Po}\big(\lambda_n\kappa_n^{-1}|_J\big)\Big)$$
$$= O\bigg(n^D w_{[2]}\,,\ m^D\bigg(n^D \wedge \frac{\log^\uparrow(n^D w_1)}{\underline{w}_1}\bigg)\big(w_1^2 \vee w_{[2]}\big)\,,\ \sqrt{n^D}\bigg(\sqrt{n^D} \wedge \frac{1}{\sqrt{\underline{w}_1}}\bigg) w_1 \breve{\beta}_n(m)\bigg)$$
for $n \to \infty$

with $\log^\uparrow(x) := 1 + \big(\log(x) \vee 0\big)$ *for* $x \ge 0$*.*

The next few results represent different attempts to simplify the above theorem by separating ξ from π_n in the various terms involved. For the first result only a slight modification in the proof of Theorem 6.3.B is necessary.

Theorem 6.3.C (L_∞-version: upper bound and convergence). *Suppose that the prerequisites of Theorem 6.3.B hold, but now replace* Π *in the mixing condition 2 by* $\widetilde{\Pi} := \xi(A_{\mathrm{int}})/|A_{\mathrm{int}}|$*, and write* $\breve{\beta}_n^{(\infty)}$*, instead of* $\breve{\beta}_n$*, for the function bounding the covariances. Furthermore set*

$$p_n^{(\infty)} := \Big\|\sup_{s\in J_n} \pi_n(s)\Big\|_{L_\infty}.$$

Then, for arbitrary $m := m(n) \in \mathbb{N}$*, we obtain*

$$d_2\Big(\mathscr{L}\big(\xi_{\pi_n}\kappa_n^{-1}|_J\big), \mathrm{Po}\big(\lambda_n\kappa_n^{-1}|_J\big)\Big)$$
$$= O\bigg(n^D (p_n^{(\infty)})^2\,,\ m^D\bigg(n^D \wedge \frac{\log^\uparrow(n^D p_n^{(\infty)})}{\underline{w}_1}\bigg)(p_n^{(\infty)})^2\,,\ \sqrt{n^D}\bigg(\sqrt{n^D} \wedge \frac{1}{\sqrt{\underline{w}_1}}\bigg) p_n^{(\infty)} \breve{\beta}_n^{(\infty)}(m)\bigg)$$
for $n \to \infty$*.*

Convergence Condition: *The right hand side goes to* 0 *if, for example,* $p_n^{(\infty)} = O(1/n^D)$ *and there is a sequence* $(m(n))_n$ *with* $m(n) = o(n)$ *such that* $\breve{\beta}_n^{(\infty)}(m(n)) = o(1)$ *as* $n \to \infty$*.*

Remark 6.3.D (Convergence towards a fixed Poisson process). If in fact ξ and π_n are such that

$$\xi_{\pi_n}\kappa_n^{-1}|_J \xrightarrow{\mathscr{D}} \mathrm{Po}(\lambda) \quad (n \to \infty)$$

for some finite measure λ on J, then we obtain by Theorem 6.2.A that, on J,

$$\int_{\kappa_n^{-1}(\cdot)} \pi_n(s)\ \xi(ds) \xrightarrow{\mathscr{D}} \lambda(\cdot) \quad (n \to \infty),$$

which implies under the convergence condition in Theorem 6.3.C that also

$$\lambda_n\kappa_n^{-1}(A) \longrightarrow \lambda(A) \quad (n \to \infty)$$

for any Borel set $A \subset J$ with $\lambda(\partial A) = 0$. Thus, by Inequality (6.3) and Theorem 6.3.C, we get an upper bound for $d_2\big(\mathscr{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathrm{Po}(\lambda)\big)$ that goes to zero under the convergence condition. Of course, the conditions are stronger than the ones for Proposition 6.2.B, but in return the upper bound is much more explicit and easier to apply.

The next result is a direct consequence of Theorem 6.3.B. This time, we try to gain a convergence rate that goes to zero under the weaker assumption on the sequence (π_n) that was used for Theorem 6.2.A, namely that $\sup_{s\in\mathbb{R}_+^D}\pi_n(s)\xrightarrow{\mathscr{D}}0$. It should be noted that the simple but rather crude estimates we are using here, are based on the assumption that the $\xi(C_{\boldsymbol{k}})$ have a generous number of moments. It is obviously by no means the best result attainable from Theorem 6.3.B.

Corollary 6.3.E (L_1-version: upper bound and convergence). *Suppose that the prerequisites of Theorem 6.3.B hold, and set*

$$p_n^{(1)}:=\sup_{1\le\boldsymbol{k}\le n}\mathbb{E}\Big(\sup_{s\in C_{\boldsymbol{k}}}\pi_n(s)\Big),\quad p_n^{(2)}:=\sup_{1\le\boldsymbol{k}\le n}\mathbb{E}\Big(\sup_{s\in C_{\boldsymbol{k}}}\pi_n(s)^2\Big).$$

Then, for arbitrary $m:=m(n)\in\mathbb{N}$, and $Q_1:=Q_1(n)$, $Q_2:=Q_2(n)\in\mathbb{Z}_+$, we obtain

$$\begin{aligned}
&d_2\Big(\mathscr{L}\big(\xi_{\pi_n}\kappa_n^{-1}|_J\big),\mathrm{Po}\big(\lambda_n\kappa_n^{-1}|_J\big)\Big)\\
&\quad=O\Bigg(n^D\Big(Q_2p_n^{(2)}+\sup_{1\le\boldsymbol{k}\le n}\mathbb{E}\big(\xi(C_{\boldsymbol{k}})^2\mathbf{1}_{\{\xi(C_{\boldsymbol{k}})^2>Q_2\}}\big)\Big),\\
&\qquad m^D\Big(n^D\wedge\frac{\log^\uparrow(n^Dw_1)}{\underline{w}_1}\Big)\Big(Q_2p_n^{(2)}+\sup_{1\le\boldsymbol{k}\le n}\mathbb{E}\big(\xi(C_{\boldsymbol{k}})^2\mathbf{1}_{\{\xi(C_{\boldsymbol{k}})^2>Q_2\}}\big)\Big),\\
&\qquad m^D\Big(n^D\wedge\frac{\log^\uparrow(n^Dw_1)}{\underline{w}_1}\Big)\Big(Q_1p_n^{(1)}+\sup_{1\le\boldsymbol{k}\le n}\mathbb{E}\big(\xi(C_{\boldsymbol{k}})\mathbf{1}_{\{\xi(C_{\boldsymbol{k}})>Q_1\}}\big)\Big)^2,\\
&\qquad\sqrt{n^D}\Big(\sqrt{n^D}\wedge\frac{1}{\sqrt{\underline{w}_1}}\Big)\Big(Q_1p_n^{(1)}+\sup_{1\le\boldsymbol{k}\le n}\mathbb{E}\big(\xi(C_{\boldsymbol{k}})\mathbf{1}_{\{\xi(C_{\boldsymbol{k}})>Q_1\}}\big)\Big)\breve{\beta}_n(m)\Bigg)\\
&\qquad\qquad\text{for } n\to\infty.
\end{aligned}$$

Convergence Condition: *Provided that $\xi(C_{\boldsymbol{k}})$ has sufficiently many moments that are bounded uniformly in $\boldsymbol{k}$, we can choose $m:=m(n)$, $Q_1:=Q_1(n)$, and $Q_2:=Q_2(n)$ in such a way that the right hand side goes to 0. For example, under the assumptions that $p_n^{(1)}=O\big(1/n^D\big)$, $p_n^{(2)}=O\big(1/n^{(1+x)D}\big)$ $(n\to\infty)$ for some $x\in(0,1]$, and $\breve{\beta}_n(t)=\breve{\beta}(t)=O(1/t^{sD})$ $(t\to\infty)$ for some $s>0$, we get convergence to zero of the upper bound if*

$$\sup_{\boldsymbol{k}\in\mathbb{N}^D}\mathbb{E}\big(\xi(C_{\boldsymbol{k}})^r\big)<\infty\text{ for some }r>2+\frac{2s+1}{sx}.$$

We now examine how certain independence properties can be exploited. The main benefit of this is a more convenient mixing condition, that takes only the point process into account, allowing the retention field to be dealt with separately.

Assumption 2′ (Mixing property). Let $\breve{\beta}^{(\mathrm{ind})}:\mathbb{R}_+\to\mathbb{R}_+$ be a decreasing function such that for every open cube $A_{\mathrm{int}}=(a,a+h\mathbf{1})$ of side length $h\in(0,1]$ and with minimal corner at $a\in\mathbb{R}_+^D$, and every surrounding set $A_{\mathrm{ext}}^{(t)}:=\mathbb{R}_+^D\setminus[a-t\mathbf{1},a+(t+h)\mathbf{1}]$ with $t\in\mathbb{R}_+$, we have that

$$\sup_{\substack{F\in\mathcal{F}_{\mathrm{ext}}^{(t)}\\ Z\in L_2(\mathcal{F}_{\mathrm{int}}),\,0\le Z\le\widetilde{\Pi}}}\big|\mathrm{cov}\big(Z,\mathbf{1}_F\big)\big|\le\breve{\beta}^{(\mathrm{ind})}(t),$$

where $\mathcal{F}_{\mathrm{int}}=\sigma\big(\xi|_{A_{\mathrm{int}}}\big)$, $\mathcal{F}_{\mathrm{ext}}^{(t)}=\sigma\big(\xi|_{A_{\mathrm{ext}}^{(t)}}\big)$, and $\widetilde{\Pi}:=\xi(A_{\mathrm{int}})/|A_{\mathrm{int}}|$.

Theorem 6.3.F (Independence of ξ and π_n: upper bound and convergence). *Suppose that the prerequisites of Theorem 6.3.B hold with Assumption 2′, instead of Assumption 2. Let ξ and π_n be independent for any $n \in \mathbb{N}$. Note that we now have*

$$\lambda_n(\cdot) = \mathbb{E}\bigg(\int_\cdot \pi_n(s)\ \xi(ds)\bigg) = \int_\cdot \mathbb{E}\pi_n(s)\ \mu_1(ds).$$

Choose $m := m(n) \in \mathbb{N}$ in such a way that $\pi_n|_{\widetilde{A}_{\mathrm{int}}}$ and $\pi_n|_{\widetilde{A}^{(m)}_{\mathrm{ext}}}$ are independent for every unit cube $\widetilde{A}_{\mathrm{int}} = (a, a+\mathbf{1})$ and every surrounding set $\widetilde{A}^{(m)}_{\mathrm{ext}} := \mathbb{R}^D_+ \setminus [a - m\mathbf{1}, a + (m+1)\mathbf{1}]$, $a \in \mathbb{R}^D_+$. Then

$$\begin{aligned} &d_2\Big(\mathscr{L}\big(\xi_{\pi_n}\kappa_n^{-1}|_J\big), \mathrm{Po}\big(\lambda_n\kappa_n^{-1}|_J\big)\Big) \\ &\quad = O\bigg(n^D p_n^{(2)}\,,\ m^D\bigg(n^D \wedge \frac{\log^{\uparrow}(n^D p_n^{(1)})}{\underline{w}_1}\bigg)p_n^{(2)}\,,\ \sqrt{n^D}\bigg(\sqrt{n^D} \wedge \frac{1}{\sqrt{\underline{w}_1}}\bigg)p_n^{(1)}\breve{\beta}^{(\mathrm{ind})}(m)\bigg) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } n \to \infty. \end{aligned}$$

Convergence Condition: *The right hand side goes to 0 if, for example, $p_n^{(1)} = O\big(1/n^D\big)$, $p_n^{(2)} = o\big(1/n^D\big)$ for $n \to \infty$, $\breve{\beta}^{(\mathrm{ind})}(t) = o(1)$ for $t \to \infty$, and there is a sequence $(m(n))_n$ with $m(n) = o\big(1/\big(n(p_n^{(2)})^{1/D}\big)\big)$ such that $\pi_n|_{\widetilde{A}_{\mathrm{int}}}$ and $\pi_n|_{\widetilde{A}^{(m(n))}_{\mathrm{ext}}}$ are independent for all sets $\widetilde{A}_{\mathrm{int}}$ and $\widetilde{A}^{(m(n))}_{\mathrm{ext}}$ of the above form.*

Proofs

A complete proof is presented only for Theorem 6.3.B. For the other statements the corresponding modifications are given.

Proof of Theorem 6.3.B. Let $\eta_n \sim \mathrm{Po}(\lambda_n)$. Our processes $\xi_{\pi_n}\kappa_n^{-1}|_J$ and $\eta_n\kappa_n^{-1}|_J$ are discretized in the following way. By Assumption 1 we may suppose that $\xi(\omega)(J_n \setminus J_n') = \eta(\omega)(J_n \setminus J_n') = 0$ for every $\omega \in \Omega$, without changing the distributions of the point processes. Then, subdivide J_n' in the domain of the contraction κ_n into the hypercubes $C_{\boldsymbol{k}}$, which were introduced in Section 6.1. Choose an arbitrary $\tilde{n} \in \mathbb{N}$, and further subdivide every $C_{\boldsymbol{k}}$ into cubes $C_{\boldsymbol{kr}} := \prod_{i=1}^D \big[k_i - 1 + (r_i - 1)/\tilde{n}, k_i - 1 + r_i/\tilde{n}\big)$ of side length $1/\tilde{n}$ for $\boldsymbol{r} = (r_1, \dots, r_D) \in \{1, 2, \dots, \tilde{n}\}^D$. The concrete "inner" and "outer" sets used for Assumption 2 are given by

$$A_{\mathrm{int}}(\boldsymbol{k}, \boldsymbol{r}) := \mathring{C}_{\boldsymbol{kr}}\,,\ A^{(t)}_{\mathrm{ext}}(\boldsymbol{k}) := \bigg(\bigcup\nolimits^{n}_{\substack{\boldsymbol{j}=1 \\ |\boldsymbol{j}-\boldsymbol{k}|>t}} C_{\boldsymbol{j}}\bigg)^{\circ}$$

for $t \in \mathbb{Z}_+$, $\boldsymbol{k} \in \{1, 2, \dots, n\}^D$, and $\boldsymbol{r} \in \{1, 2, \dots, \tilde{n}\}^D$, where $\mathring{C}$ denotes the interior of $C \subset \mathbb{R}^D$. We denote by $\alpha_{\boldsymbol{kr}}$ the center of $C_{\boldsymbol{kr}}$, and by $\gamma_{\boldsymbol{kr}} := \kappa_n(\alpha_{\boldsymbol{kr}})$ the center of the contracted hypercube $\kappa_n(C_{\boldsymbol{kr}})$. Furthermore set

$$I_{\boldsymbol{kr}} := 1_{\{\xi_{\pi_n}(C_{\boldsymbol{kr}}) \geq 1\}},\quad q_{\boldsymbol{kr}} := \mathbb{E}I_{\boldsymbol{kr}},$$

and let

$$Y_{\boldsymbol{kr}} \sim \mathrm{Po}(q_{\boldsymbol{kr}}),\ \boldsymbol{k} \in \{1, 2, \dots, n\}^D,\ \boldsymbol{r} \in \{1, 2, \dots, \tilde{n}\}^D,$$

be independent. Construct the discretized processes (note that, for the Poisson process, our discretization is only "in distribution") as

$$\Xi_n := \sum_{\boldsymbol{k}=1}^{n}\sum_{\boldsymbol{r}=1}^{\tilde{n}} I_{\boldsymbol{kr}}\delta_{\alpha_{\boldsymbol{kr}}}, \quad \mathrm{H}_n := \sum_{\boldsymbol{k}=1}^{n}\sum_{\boldsymbol{r}=1}^{\tilde{n}} Y_{\boldsymbol{kr}}\delta_{\alpha_{\boldsymbol{kr}}},$$

and let

$$\tilde{\mu} := \tilde{\mu}_n := \sum_{\boldsymbol{k}=1}^{n}\sum_{\boldsymbol{r}=1}^{\tilde{n}} q_{\boldsymbol{kr}}\delta_{\alpha_{\boldsymbol{kr}}} = \mathbb{E}\Xi_n = \mathbb{E}\mathrm{H}_n.$$

Note that the $\boldsymbol{r}$-sum over $q_{\boldsymbol{kr}}$ can be estimated from above as

$$\begin{aligned}\sum_{\boldsymbol{r}=1}^{\tilde{n}} q_{\boldsymbol{kr}} \le \mathbb{E}\xi_{\pi_n}(C_{\boldsymbol{k}}) &= \mathbb{E}\Big(\mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{k}}) \;\big|\; \xi, \pi_n\big)\Big)\\ &= \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})} X_i \;\middle|\; \xi, \pi_n\right)\right)\\ &= \mathbb{E}\left(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})} \pi_n(S_i)\right) \le w_1, \end{aligned} \tag{6.4}$$

where we used the "tacit numbering of points" and the "tacit definition of the retention decisions X_i" as announced in Remark 6.1.A. In an analogous way, the same sum can be estimated from below as

$$\begin{aligned}\sum_{\boldsymbol{r}=1}^{\tilde{n}} q_{\boldsymbol{kr}} &\ge \mathbb{E}\Big(\mathbb{P}\Big[\textstyle\sum_{i=1}^{\xi(C_{\boldsymbol{k}})} X_i \ge 1 \;\Big|\; \xi, \pi_n\Big]\Big)\\ &\ge \mathbb{E}\Big(1_{\{\xi(C_{\boldsymbol{k}})\ge 1\}} \sup_{1\le i\le \xi(C_{\boldsymbol{k}})} \pi_n(S_i)\Big) \ge \underline{w}_1.\end{aligned}$$

The initial distance is split up as follows:

$$\begin{aligned} d_2\big(\mathscr{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathscr{L}(\eta_n\kappa_n^{-1}|_J)\big) \le\; & d_2\big(\mathscr{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathscr{L}(\Xi_n\kappa_n^{-1})\big)\\ &+ d_2\big(\mathscr{L}(\Xi_n\kappa_n^{-1}), \mathscr{L}(\mathrm{H}_n\kappa_n^{-1})\big)\\ &+ d_2\big(\mathscr{L}(\mathrm{H}_n\kappa_n^{-1}), \mathscr{L}(\eta_n\kappa_n^{-1}|_J)\big). \end{aligned} \tag{6.5}$$

We first attend to the discretization errors, which are represented by the first and third terms on the right hand side. We have, by Theorem 3.2.A(i) and Lemma 3.1.A(ii),

$$\begin{aligned} &d_2\big(\mathscr{L}(\xi_{\pi_n}\kappa_n^{-1}|_J), \mathscr{L}(\Xi_n\kappa_n^{-1})\big)\\ &\quad\le \mathbb{E}\Big(d_1\big(\xi_{\pi_n}\kappa_n^{-1}|_J, \Xi_n\kappa_n^{-1}\big)1_{\{\xi_{\pi_n}(J_n)=\Xi_n(J_n)\}}\Big) + \mathbb{P}\big[\xi_{\pi_n}(J_n) \ne \Xi_n(J_n)\big]\\ &\quad= \mathbb{E}\Big(d_1\big(\xi_{\pi_n}\kappa_n^{-1}|_J, \Xi_n\kappa_n^{-1}\big)1_{\{\xi_{\pi_n}(C_{\boldsymbol{kr}})=\Xi_n(C_{\boldsymbol{kr}})\,\forall \boldsymbol{k},\boldsymbol{r}\}}\Big) + \mathbb{P}\big[\textstyle\bigcup_{\boldsymbol{k},\boldsymbol{r}}\{\xi_{\pi_n}(C_{\boldsymbol{kr}}) \ge 2\}\big]\\ &\quad\le \sqrt{D}/(2n\tilde{n}) + \sum_{\boldsymbol{k}=1}^{n}\sum_{\boldsymbol{r}=1}^{\tilde{n}} \mathbb{P}[\xi_{\pi_n}(C_{\boldsymbol{kr}}) \ge 2]. \end{aligned}$$

The first summand in the last line is obtained because, for estimating the d_1-distance on $\{\xi_{\pi_n}(C_{\boldsymbol{kr}}) = \Xi_n(C_{\boldsymbol{kr}})\ \forall \boldsymbol{k}, \boldsymbol{r}\}$, we can pair every point of $\xi_{\pi_n}\kappa_n^{-1}|_J$ with the center of that cube $\kappa_n(C_{\boldsymbol{kr}})$ in which it lies; this center is at most $\sqrt{D}/(2n\tilde{n})$ (half a body diagonal of $\kappa_n(C_{\boldsymbol{kr}})$) apart and contains a point of $\Xi_n\kappa_n^{-1}$. A similar argument is used to obtain the term $\sqrt{D}/(2n\tilde{n})$ in the fourth line of the next formula, with the difference that now there might be more than one point pairing in each $\kappa_n(C_{\boldsymbol{kr}})$. Thus

$$\begin{aligned}
&d_2\big(\mathscr{L}(\eta_n\kappa_n^{-1}|_J), \mathscr{L}(\mathrm{H}_n\kappa_n^{-1})\big)\\
&\qquad\le d_2\Big(\mathscr{L}\big(\eta_n\kappa_n^{-1}|_J\big), \mathscr{L}\big(\textstyle\sum_{\boldsymbol{k},\boldsymbol{r}}\eta_n(C_{\boldsymbol{kr}})\delta_{\gamma_{\boldsymbol{kr}}}\big)\Big)\\
&\qquad\quad + d_2\Big(\mathscr{L}\big(\textstyle\sum_{\boldsymbol{k},\boldsymbol{r}}\eta_n(C_{\boldsymbol{kr}})\delta_{\gamma_{\boldsymbol{kr}}}\big), \mathscr{L}\big(\textstyle\sum_{\boldsymbol{k},\boldsymbol{r}}Y_{\boldsymbol{kr}}\delta_{\gamma_{\boldsymbol{kr}}}\big)\Big)\\
&\qquad\le \sqrt{D}/(2n\tilde{n}) + d_{TV}\Big(\mathscr{L}\big(\textstyle\sum_{\boldsymbol{k},\boldsymbol{r}}\eta_n(C_{\boldsymbol{kr}})\delta_{\gamma_{\boldsymbol{kr}}}\big), \mathscr{L}\big(\textstyle\sum_{\boldsymbol{k},\boldsymbol{r}}Y_{\boldsymbol{kr}}\delta_{\gamma_{\boldsymbol{kr}}}\big)\Big)\\
&\qquad\le \sqrt{D}/(2n\tilde{n}) + \sum_{\boldsymbol{k},\boldsymbol{r}} d_{TV}\big(\mathscr{L}(\eta_n(C_{\boldsymbol{kr}})), \mathscr{L}(Y_{\boldsymbol{kr}})\big)\\
&\qquad\le \sqrt{D}/(2n\tilde{n}) + \sum_{\boldsymbol{k},\boldsymbol{r}} \Big|\mathbb{E}\xi_{\pi_n}(C_{\boldsymbol{kr}}) - \mathbb{P}[\xi_{\pi_n}(C_{\boldsymbol{kr}}) \ge 1]\Big|\\
&\qquad= \sqrt{D}/(2n\tilde{n}) + \sum_{\boldsymbol{k},\boldsymbol{r}} \mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{kr}})1_{\{\xi_{\pi_n}(C_{\boldsymbol{kr}})\ge 2\}}\big) - \sum_{\boldsymbol{k},\boldsymbol{r}} \mathbb{P}\big[\xi_{\pi_n}(C_{\boldsymbol{kr}}) \ge 2\big],
\end{aligned}$$

where the various estimation steps for the second summand follow along the lines of Inequality (5.6). Since the sum over $\boldsymbol{r}$ in the second term can be estimated as

$$\begin{aligned}
\sum_{\boldsymbol{r}=1}^{\tilde{n}} \mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{kr}})1_{\{\xi_{\pi_n}(C_{\boldsymbol{kr}})\ge 2\}}\big) &\le \mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{k}})1_{\{\xi_{\pi_n}(C_{\boldsymbol{k}})\ge 2\}}\big)\\
&\le \mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{k}})(\xi_{\pi_n}(C_{\boldsymbol{k}}) - 1)\big)\\
&= \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})}\sum_{\substack{j=1\\ j\ne i}}^{\xi(C_{\boldsymbol{k}})} X_iX_j \;\middle|\; \xi, \pi_n\right)\right)\\
&= \mathbb{E}\left(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})}\sum_{\substack{j=1\\ j\ne i}}^{\xi(C_{\boldsymbol{k}})} \pi_n(S_i)\pi_n(S_j)\right)\\
&\le \mathbb{E}\left(\xi(C_{\boldsymbol{k}})\big(\xi(C_{\boldsymbol{k}}) - 1\big)\Big(\sup_{1\le i\le \xi(C_{\boldsymbol{k}})} \pi_n(S_i, \cdot)\Big)^2\right)\\
&\le w_{[2]},
\end{aligned} \tag{6.6}$$

we obtain, as an overall bound for the discretization error terms,

$$\sqrt{D}/(n\tilde{n}) + n^D w_{[2]}. \tag{6.7}$$

Next we consider the left over term in (6.5), which is estimated by application of the local Stein Theorem 4.3.A. We choose $m := m(n) \in \mathbb{N}$ arbitrarily, and set in the notation

of Section 4.3

$$\begin{aligned}\Theta &= \{1,2,\dots,n\}^D \times \{1,2,\dots,\tilde{n}\}^D\\ \Theta_{\boldsymbol{kr}} &= \Theta \setminus \{(\boldsymbol{k},\boldsymbol{r})\}\\ \Theta^s_{\boldsymbol{kr}} &= \{(\boldsymbol{l},\boldsymbol{s}) \in \Theta_{\boldsymbol{kr}}\,;\ |\boldsymbol{l}-\boldsymbol{k}| \le m\}\\ \Theta^w_{\boldsymbol{kr}} &= \{(\boldsymbol{l},\boldsymbol{s}) \in \Theta_{\boldsymbol{kr}}\,;\ |\boldsymbol{l}-\boldsymbol{k}| > m\}.\end{aligned}$$

Since $\mathbb{E}\Xi_n = \mathbb{E}\mathrm{H}_n = \tilde{\mu}$, we can apply Theorem 4.3.A and obtain (once more in the notation of Section 4.3):

$$d_2\big(\mathscr{L}(\Xi_n\kappa_n^{-1}), \mathscr{L}(\mathrm{H}_n\kappa_n^{-1})\big) \le M_1(\tilde{\mu}) \sum_{\boldsymbol{k},\boldsymbol{r}} \big(q_{\boldsymbol{kr}}\mathbb{E}\widetilde{Z}_{\boldsymbol{kr}} + \mathbb{E}(I_{\boldsymbol{kr}}Z_{\boldsymbol{kr}})\big) + M_2(\tilde{\mu}) \sum_{\boldsymbol{k},\boldsymbol{r}} e_{\boldsymbol{kr}}.$$

Further estimation of the various terms yields

$$M_1(\tilde{\mu}) \le 1 \wedge \left[\frac{2}{n^D \underline{w}_1}\Big(1 + 2\log^+\Big(\frac{n^D w_1}{2}\Big)\Big)\right]; \tag{6.8}$$

$$M_2(\tilde{\mu}) \le 1 \wedge \frac{1.65}{\sqrt{n^D \underline{w}_1}}; \tag{6.9}$$

$$\sum_{\boldsymbol{k},\boldsymbol{r}} q_{\boldsymbol{kr}}\mathbb{E}\widetilde{Z}_{\boldsymbol{kr}} = \sum_{\boldsymbol{k}=1}^{n} \sum_{\substack{\boldsymbol{l}=1\\ |\boldsymbol{l}-\boldsymbol{k}|\le m}}^{n} \Big(\sum_{\boldsymbol{r}=1}^{\tilde{n}} q_{\boldsymbol{kr}}\Big)\Big(\sum_{\boldsymbol{s}=1}^{\tilde{n}} q_{\boldsymbol{ls}}\Big) \le (2m+1)^D n^D w_1^2; \tag{6.10}$$

and, using part of Inequality (6.6) for the second line,

$$\begin{aligned}\sum_{\boldsymbol{k},\boldsymbol{r}} \mathbb{E}(I_{\boldsymbol{kr}}Z_{\boldsymbol{kr}}) &\le \sum_{\boldsymbol{k}=1}^{n}\Bigg(\sum_{\substack{\boldsymbol{l}=1\\ 1\le|\boldsymbol{l}-\boldsymbol{k}|\le m}}^{n} \mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{k}})\,\xi_{\pi_n}(C_{\boldsymbol{l}})\big) \;+\; \mathbb{E}\big(\xi_{\pi_n}(C_{\boldsymbol{k}})(\xi_{\pi_n}(C_{\boldsymbol{k}})-1)\big)\Bigg)\\ &\le \sum_{\substack{\boldsymbol{k},\boldsymbol{l}=1\\ 1\le|\boldsymbol{l}-\boldsymbol{k}|\le m}}^{n} \mathbb{E}\Bigg(\mathbb{E}\Bigg(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})}\sum_{j=\xi(C_{\boldsymbol{k}})+1}^{\xi(C_{\boldsymbol{k}})+\xi(C_{\boldsymbol{l}})} X_iX_j \;\Big|\; \xi,\pi_n\Bigg)\Bigg) \;+\; n^D w_{[2]}\\ &= \sum_{\substack{\boldsymbol{k},\boldsymbol{l}=1\\ 1\le|\boldsymbol{l}-\boldsymbol{k}|\le m}}^{n} \mathbb{E}\Bigg(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})}\sum_{j=\xi(C_{\boldsymbol{k}})+1}^{\xi(C_{\boldsymbol{k}})+\xi(C_{\boldsymbol{l}})} \pi_n(S_i)\pi_n(S_j)\Bigg) \;+\; n^D w_{[2]}\\ &\le (2m+1)^D n^D w_{[2]},\end{aligned} \tag{6.11}$$

where our numbering in Inequality (6.11) is such that the points $S_1,\dots,S_{\xi(C_{\boldsymbol{k}})}$ lie in $C_{\boldsymbol{k}}$ and the points $S_{\xi(C_{\boldsymbol{k}})+1},\dots,S_{\xi(C_{\boldsymbol{k}})+\xi(C_{\boldsymbol{l}})}$ in $C_{\boldsymbol{l}}$.

A little more work is needed to estimate $e_{\boldsymbol{kr}}$. First, note that by Assumption 1, the probability that any points of ξ lie on the grid $G := \bigcup_{\boldsymbol{kr}} \partial C_{\boldsymbol{kr}}$ is zero. Since we are only interested in distributional properties of ξ_π, we may therefore assume without loss of generality that $\xi(\omega)(G) = 0$ for all $\omega \in \Omega$. For any set U, denote its power set by $\mathcal{P}(U)$, and write

$$\mathcal{F}^w_{\boldsymbol{k}} := \mathcal{P}\big(\{0,1\}^{\Theta^w_{\boldsymbol{kr}}}\big) \text{ and } W_{\boldsymbol{k}} := (I_{\boldsymbol{ls}})_{(\boldsymbol{l},\boldsymbol{s})\in\Theta^w_{\boldsymbol{kr}}}$$

(note that $\Theta^w_{\boldsymbol{kr}}$ does not depend on $\boldsymbol{r}$). If $\mathbb{E}\big(\xi(C_{\boldsymbol{kr}})\sup_{s\in C_{\boldsymbol{kr}}}\pi_n(s)\big)$ is zero, it is evident that $e_{\boldsymbol{kr}}=0$, so assume that the expectation is non-zero. We use the "formula of total covariance", that is, the relation

$$\operatorname{cov}(X,Y)=\mathbb{E}\big(\operatorname{cov}(X,Y\,|\,Z)\big)+\operatorname{cov}\big(\mathbb{E}(X\,|\,Z),\mathbb{E}(Y\,|\,Z)\big)$$

for random variables $X,Y\in L_2$ and an arbitrary random element Z, along with the conditional independence of the retention decisions, to obtain

$$\begin{aligned} e_{\boldsymbol{kr}}/2 &= \max_{B\in\mathcal{F}^w_{\boldsymbol{k}}}\Big|\operatorname{cov}\big(1_{\{\xi_{\pi_n}(C_{\boldsymbol{kr}})\geq 1\}}\,,\;1_{\{W_{\boldsymbol{k}}\in B\}}\big)\Big| \\ &= \max_{B\in\mathcal{F}^w_{\boldsymbol{k}}}\Big|\operatorname{cov}\Big(\mathbb{P}\big[\xi_{\pi_n}(C_{\boldsymbol{kr}})\geq 1\;\big|\;\xi,\pi_n\big]\,,\;\mathbb{P}\big[W_{\boldsymbol{k}}\in B\;\big|\;\xi,\pi_n\big]\Big)\Big|. \end{aligned} \tag{6.12}$$

Since no realization of ξ has any points in G, the arguments of the covariance are conditional probabilities of the form $\mathbb{P}[\xi_{\pi_n}\in N\,|\,\xi,\pi_n]$ with

$$N\in\mathcal{N}(A):=\sigma\big(\big\{\{\varrho\in\mathfrak{N};\varrho(\tilde{A})=l\}\,;\;\tilde{A}\in\mathcal{B}_A,l\in\mathbb{Z}_+\big\}\big)\subset\mathcal{N},$$

where $A=A_{\mathrm{int}}(\boldsymbol{k},\boldsymbol{r})$ and $A=A^{(m)}_{\mathrm{ext}}(\boldsymbol{k})$, respectively. We then may condition as well only on ξ and π_n restricted to the corresponding set A (the proof of this is rather technical and has therefore been placed in Appendix A.4.2). Hence

$$\begin{aligned} \frac{1}{2}\sum_{\boldsymbol{r}=1}^{\tilde{n}} e_{\boldsymbol{kr}} &= \sum_{\boldsymbol{r}=1}^{\tilde{n}}\max_{B\in\mathcal{F}^w_{\boldsymbol{k}}}\Big|\operatorname{cov}\Big(\mathbb{P}\big[\xi_{\pi_n}(C_{\boldsymbol{kr}})\geq 1\;\big|\;\xi|_{A_{\mathrm{int}}(\boldsymbol{k},\boldsymbol{r})},\pi_n|_{A_{\mathrm{int}}(\boldsymbol{k},\boldsymbol{r})}\big], \\ &\qquad\qquad \mathbb{P}\big[W_{\boldsymbol{k}}\in B\;\big|\;\xi|_{A^{(m)}_{\mathrm{ext}}(\boldsymbol{k})},\pi_n|_{A^{(m)}_{\mathrm{ext}}(\boldsymbol{k})}\big]\Big)\Big| \\ &\leq w_1\,\breve{\beta}_n(m), \end{aligned} \tag{6.13}$$

where, for the inequality, the factor $\mathbb{E}\big(\xi(C_{\boldsymbol{kr}})\sup_{s\in C_{\boldsymbol{kr}}}\pi_n(s)\big)$ was extracted from the first argument of the covariance. Note that in the second argument we do not have an indicator as required, but a general $[0,1]$-valued random variable. The upper bound from Assumption 2 still holds, as can be seen from the proof of Equation $(1')$ in Doukhan (1994), Section 1.1.

We assemble the different parts from Result (6.7), Inequalities (6.8) to (6.11), and Inequality (6.13), and let $\tilde{n}$ go to infinity to obtain the overall estimate

$$\begin{aligned} &d_2\Big(\mathscr{L}\big(\xi_{\pi_n}\kappa_n^{-1}|_J\big),\mathscr{L}\big(\eta_n\kappa_n^{-1}|_J\big)\Big) \\ &\quad\leq\; n^D w_{[2]}+\left(1\wedge\left[\frac{2}{n^D\underline{w}_1}\Big(1+2\log^+\Big(\frac{n^D w_1}{2}\Big)\Big)\right]\right)(2m+1)^D n^D(w_1^2+w_{[2]}) \\ &\qquad+\left(1\wedge\frac{1.65}{\sqrt{n^D\underline{w}_1}}\right)2n^D w_1\breve{\beta}_n(m), \end{aligned}$$

which is of the required order for $n\to\infty$. □

Proof of Theorem 6.3.C. Use in the bound of Theorem 6.3.B the estimates $w_1\leq\bar{h}_1 p_n^{(\infty)}$ and $w_{[2]}\leq\bar{h}_2(p_n^{(\infty)})^2$, and modify the estimation of the $e_{\boldsymbol{kr}}$ by extracting $p_n^{(\infty)}|\mathrm{C}_{\boldsymbol{kr}}|$ instead

of $\mathbb{E}\big(\xi(C_{\boldsymbol{kr}})\sup_{s\in C_{\boldsymbol{kr}}}\pi_n(s)\big)$ in Inequality (6.13), so that we get

$$\frac{1}{2}\sum_{\boldsymbol{r}=1}^{\tilde{n}} e_{\boldsymbol{kr}} \le p_n^{(\infty)}\breve{\beta}_n^{(\infty)}(m).$$

□

Proof of Corollary 6.3.E. In the bound of Theorem 6.3.B, use the estimates

$$w_1 \le Q_1 p_n^{(1)} + \sup_{1\le\boldsymbol{k}\le n}\mathbb{E}\big(\xi(C_{\boldsymbol{k}})1_{\{\xi(C_{\boldsymbol{k}})>Q_1\}}\big),\ \ w_{[2]} \le Q_2 p_n^{(2)} + \sup_{1\le\boldsymbol{k}\le n}\mathbb{E}\big(\xi(C_{\boldsymbol{k}})^2 1_{\{\xi(C_{\boldsymbol{k}})^2>Q_2\}}\big),$$

from which we obtain the required order. For the convergence condition, set $m(n) := n^y$, $Q_1^{(\varepsilon)}(n) := n^{(sy-\varepsilon)D}$ and $Q_2^{(\varepsilon)}(n) := n^{(2s(x+1)y-\varepsilon)D}$, where $y := x\big/\big(2s(x+1)+1\big)$ and $\varepsilon>0$ is arbitrary. Use that $\mathbb{E}(Y1_{\{Y>Q\}}) = o\big((1/Q)^{r-1}\big)$ for any non-negative random variable Y with finite r-th moment, in order to show that for any $r > 2+(2s+1)/(sx)$, there is still a choice of $\varepsilon>0$ such that the upper bound in Corollary 6.3.E goes to zero. □

Proof of Theorem 6.3.F. Use in the bound of Theorem 6.3.B the estimates $w_1 \le \bar{h}_1 p_n^{(1)}$ and $w_{[2]} \le \bar{h}_2 p_n^{(2)}$, and modify the estimation of the $e_{\boldsymbol{kr}}$ in the following way: for better readability, suppress the indices $\boldsymbol{k}$ and $\boldsymbol{r}$ in the expressions $A_{\mathrm{int}}(\boldsymbol{k},\boldsymbol{r})$, $A_{\mathrm{ext}}^{(m)}(\boldsymbol{k})$, and write $f\big(\xi|_{A_{\mathrm{int}}},\pi_n|_{A_{\mathrm{int}}}\big)$ and $g\big(\xi|_{A_{\mathrm{ext}}^{(m)}},\pi_n|_{A_{\mathrm{ext}}^{(m)}}\big) = g_B\big(\xi|_{A_{\mathrm{ext}}^{(m)}},\pi_n|_{A_{\mathrm{ext}}^{(m)}}\big)$ for the conditional probabilities in (6.13), such that

$$e_{\boldsymbol{kr}}/2 = \max_{B\in\mathcal{F}_{\boldsymbol{k}}^{w}}\Big|\mathrm{cov}\Big(f\big(\xi|_{A_{\mathrm{int}}},\pi_n|_{A_{\mathrm{int}}}\big)\,,\ g_B\big(\xi|_{A_{\mathrm{ext}}^{(m)}},\pi_n|_{A_{\mathrm{ext}}^{(m)}}\big)\Big)\Big|.$$

Then, because of the independence of $\big(\xi,\pi_n|_{A_{\mathrm{int}}},\pi_n|_{A_{\mathrm{ext}}^{(m)}}\big)$,

$$\begin{aligned}
&\Big|\mathrm{cov}\Big(f\big(\xi|_{A_{\mathrm{int}}},\pi_n|_{A_{\mathrm{int}}}\big)\,,\ g\big(\xi|_{A_{\mathrm{ext}}^{(m)}},\pi_n|_{A_{\mathrm{ext}}^{(m)}}\big)\Big)\Big| \\
&\quad\le \mathbb{E}\Big|\mathrm{cov}\Big(f\big(\xi|_{A_{\mathrm{int}}},\pi_n|_{A_{\mathrm{int}}}\big)\,,\ g\big(\xi|_{A_{\mathrm{ext}}^{(m)}},\pi_n|_{A_{\mathrm{ext}}^{(m)}}\big)\ \Big|\ \pi_n|_{A_{\mathrm{int}}},\pi_n|_{A_{\mathrm{ext}}^{(m)}}\Big)\Big| \\
&\quad\le |A_{\mathrm{int}}|\,\mathbb{E}\Bigg(\sup_{s\in C_{\boldsymbol{k}}}\pi_n(s)\,\Big|\mathrm{cov}\Big(\tfrac{f(\xi|_{A_{\mathrm{int}}},\pi_n|_{A_{\mathrm{int}}})}{|A_{\mathrm{int}}|\sup_{s\in A_{\mathrm{int}}}\pi_n(s)}\,,\ g\big(\xi|_{A_{\mathrm{ext}}^{(m)}},\pi_n|_{A_{\mathrm{ext}}^{(m)}}\big)\ \Big|\ \pi_n|_{A_{\mathrm{int}}},\pi_n|_{A_{\mathrm{ext}}^{(m)}}\Big)\Big|\Bigg) \\
&\quad\le |A_{\mathrm{int}}|\,p_n^{(1)}\breve{\beta}^{(\mathrm{ind})}(m),
\end{aligned}$$

hence

$$\frac{1}{2}\sum_{\boldsymbol{r}=1}^{\tilde{n}} e_{\boldsymbol{kr}} \le p_n^{(1)}\breve{\beta}^{(\mathrm{ind})}(m),$$

which yields the desired result. □

6.4 An alternative thinning definition

In this section, we consider a different thinning concept, where in place of a random retention field, we have deterministic retention kernels by which we directly model dependences between retention decisions. This will lead us, by means of the same method

of proof as before, to a theorem that is similar to Theorem 6.3.B, less appealing from a theoretical and a typographical point of view, but sometimes more intuitively applied, because it permits us to look at the situation from a different angle: rather than thinking of a thinning as the result of point deletions according to a (potentially inscrutable) random environment, we now understand it as the result of point deletions according to (potentially more transparent) point interactions.

We first introduce the new thinning concept, which is based on the following definition.

Definition (Admissible sequence of retention kernels). For any $u \in \mathbb{N}$, let $E_u := \{(s_1, s_2, \ldots, s_u; \sigma) \in (\mathbb{R}_+^D)^u \times \mathfrak{N};\ \sum_{i=1}^u \delta_{s_i} \leq \sigma\}$, equipped with the trace σ-field of $(\mathcal{B}_+^D)^u \otimes \mathcal{N}$, and denote by $E_u(\sigma) := \{\boldsymbol{s} \in (\mathbb{R}_+^D)^u; (\boldsymbol{s}; \sigma) \in E_u\}$ the section of E_u at $\sigma \in \mathfrak{N}$. Let Q_u be a probability kernel from E_u to $\{0,1\}^u$. We call $(Q_u)_{u\in\mathbb{N}}$ an *admissible sequence of retention kernels* if

(a) the Q_u are "simultaneously symmetrical" in the sense that for any permutation τ on $\{1, \ldots, u\}$ and corresponding transformations $T_1 : \big(\mathbb{R}_+^D\big)^u \to \big(\mathbb{R}_+^D\big)^u$, $(s_1, \ldots, s_u) \mapsto (s_{\tau(1)}, \ldots, s_{\tau(u)})$ and $T_2 : \{0,1\}^u \to \{0,1\}^u$, $(e_1, \ldots, e_u) \mapsto (e_{\tau(1)}, \ldots, e_{\tau(u)})$ we have

$$Q_u\big((T_1\boldsymbol{s}; \sigma), T_2(A)\big) = Q_u\big((\boldsymbol{s}; \sigma), A\big)$$

for every $(\boldsymbol{s}, \sigma) \in E_u$ and every $A \subset \{0,1\}^u$;

(b) the following compatibility condition holds between the Q_u: let $u \geq 2$ be arbitrary; then we have

$$Q_u\big((s_1, \ldots, s_{u-1}, s_u; \sigma), A \times \{0,1\}\big) = Q_{u-1}\big((s_1, \ldots, s_{u-1}; \sigma), A\big)$$

for every $(s_1, \ldots, s_{u-1}, s_u; \sigma) \in E_u$ and every $A \subset \{0,1\}^{u-1}$.

Now let ξ be a point process on $\mathbb{R}_+^D$. To simplify the presentation we assume, for this whole section, that all realizations of ξ have infinitely many points in $\mathbb{R}_+^D$. Furthermore, let $Q = (Q_u)_{u\in\mathbb{N}}$ be an admissible sequence of retention kernels. In analogy with Section 6.1, the Q-thinning of ξ can now be defined as follows.

Definition ("Q-thinning"). First, assume that $\xi = \sigma = \sum_{i=1}^\infty \delta_{s_i}$ is non-random, and define a Q-thinning of ξ in this case as $\xi_Q := \sum_{i=1}^\infty X_i \delta_{s_i}$, where the X_i are indicator random variables whose joint distribution is given by the fidi-distributions

$$\mathscr{L}(X_1, \ldots, X_u) = Q_u\big((s_1, \ldots, s_u; \sigma), \cdot\big)$$

for every $u \in \mathbb{N}$. It is easy to see that, due to Properties (a) and (b) from above, ξ_Q has a distribution P_σ that is well-defined and does not depend on the enumeration of σ. For general ξ, define then a *Q-thinning* ξ_Q by randomization, as in Section 6.1.

Remark 6.4.A. Let $f_1, f_2, \ldots$ be $\mathcal{N}$-$\mathcal{B}_+^D$-measurable functions such that $\sigma = \sum_{i=1}^{\sigma(\mathbb{R}_+^D)} \delta_{f_i(\sigma)}$ for every $\sigma \in \mathfrak{N}$. Such functions exist by Corollary 2.2.D. With any such sequence of functions it is enough to define $Q_u\big((f_1(\sigma), \ldots, f_u(\sigma); \sigma), A\big)$ for every $\sigma \in \mathfrak{N}$, every $A \subset \{0,1\}^u$, and every $u \in \mathbb{N}$, in such a way that Properties (a) and (b) from the thinning definition are satisfied, and the above term is a $\mathcal{N}$-measurable mapping in σ and a probability measure in A. There is then a unique admissible sequence $(\widetilde{Q}_u)_{u\in\mathbb{N}}$ of retention kernels such that $\widetilde{Q}_u$ extends Q_u to the whole of E_u for every $u \in \mathbb{N}$. A short proof of this statement is the topic of Appendix A.4.3.

Remark 6.4.B. The new Q-thinning concept generalizes the thinning concept from Section 6.1. That is to say, for any combination of a point process ξ and a locally evaluable random field π, the thinning ξ_π can be modeled as a Q-thinning ξ_Q. As in Remark 6.4.A, let $f_1, f_2, \ldots$ be $\mathcal{N}$-$\mathcal{B}_+^D$-measurable functions with $\sigma = \sum_i \delta_{f_i(\sigma)}$ for every $\sigma \in \mathfrak{N}$. Define then for any $u \in \mathbb{N}$, and for $e_1, \ldots, e_u \in \{0,1\}$,

$$Q_u\Big(\big(f_1(\sigma), \ldots, f_u(\sigma); \sigma\big), \big\{(e_1, \ldots, e_u)\big\}\Big) := \mathbb{E}\bigg(\prod_{i=1}^{u} \big(\pi(f_i(\xi))\big)^{e_i} \big(1 - \pi(f_i(\xi))\big)^{1-e_i} \;\Big|\; \xi = \sigma\bigg) \tag{6.14}$$

for almost every σ, which, upon adaptation on a $\mathbb{P}\xi^{-1}$-null set, yields by Remark 6.4.A, a well-defined sequence of retention kernels. It follows then from (6.14) that for every $N \in \mathcal{N}$

$$\mathbb{P}[\xi_Q \in N \,|\, \xi] = \mathbb{E}\big(\mathbb{P}[\xi_\pi \in N \,|\, \xi, \pi] \;\big|\; \xi\big) = \mathbb{P}[\xi_\pi \in N \,|\, \xi] \quad \text{a.s.},$$

and hence that ξ_Q has the same distribution as ξ_π.

One can prove a theorem corresponding to Theorem 6.3.B, which now relies on separate control of a mixing coefficient with respect to ξ alone and of the conditional covariances between functions of retention decisions. This seems intuitively more appealing, but is also quite a bit more inconvenient to formulate than the more abstract way via the σ-fields $\mathcal{F}_{\text{int}}^{(n)}$ and $\mathcal{F}_{\text{ext}}^{(n,t)}$ used for Theorem 6.3.B. To keep things reasonably neat, we only state a special case here, which is basically the analog of Theorem 6.3.C, the L_∞-version.

We first present the additional assumptions we need. Let $Q^{(n)} := \big(Q_u^{(n)}\big)_{u \in \mathbb{N}}$ for each $n \in \mathbb{N}$ be an admissible sequence of retention kernels, and set

$$\bar{p}_n := \operatorname{ess\,sup}\bigg(\sup_{s \in E_1(\xi) \cap J_n} Q_1^{(n)}\big((s;\xi), \{1\}\big)\bigg) \in [0,1],$$

$$\underline{p}_n := \operatorname{ess\,inf}\bigg(\inf_{s \in E_1(\xi) \cap J_n} Q_1^{(n)}\big((s;\xi), \{1\}\big)\bigg) \in [0,1],$$

where we define the supremum over the empty set as 0 and the infimum as 1. Note that the role of $\pi_n(\omega, s)$ is now always taken by $Q_1^{(n)}\big((s;\xi(\omega)), \{1\}\big)$ and that $E_1(\sigma)$ is just the set of all points of σ. The additional assumptions are as follows.

Assumption 2″ (Mixing property of $\boldsymbol{\xi}$). Let $\breve{\beta}^{(\infty)} : \mathbb{R}_+ \to \mathbb{R}_+$ be a decreasing function such that for any cubes $A_{\text{int}} = (a, a + h\mathbf{1})$, $A_{\text{int}}^{(\tilde{t})} = \big(a - \tilde{t}\mathbf{1}, a + (\tilde{t} + h)\mathbf{1}\big)$ and surrounding sets $A_{\text{ext}}^{(\tilde{t},t)} := \mathbb{R}_+^D \setminus \big[a - (\tilde{t} + t)\mathbf{1}, a + (\tilde{t} + t + h)\mathbf{1}\big]$ with $h \in (0,1]$, $a \in \mathbb{R}_+^D$, and $\tilde{t}, t \in \mathbb{R}_+$, we have that

$$\sup_{\substack{F \in \mathcal{F}_{\text{ext}}^{(\tilde{t},t)} \\ Z \in L_2(\mathcal{F}_{\text{int}}^{(\tilde{t})}),\, 0 \le Z \le \widetilde{\Pi}}} \big|\operatorname{cov}\big(Z, 1_F\big)\big| \le \breve{\beta}^{(\infty)}(t),$$

where $\mathcal{F}_{\text{int}}^{(\tilde{t})} := \sigma\big(\xi|_{A_{\text{int}}^{(\tilde{t})}}\big)$, $\mathcal{F}_{\text{ext}}^{(\tilde{t},t)} := \sigma\big(\xi|_{A_{\text{ext}}^{(\tilde{t},t)}}\big)$, and $\widetilde{\Pi} := \xi(A_{\text{int}})/|A_{\text{int}}|$.

Assumption 3 (Local functional dependence of $Q_u^{(n)}(\cdot, A)$ on σ). For each $n \in \mathbb{N}$ there is a $\tilde{t}_n \in \mathbb{R}_+$, such that $Q_u^{(n)}(\cdot, A)$ is only a function of $\big(s_1, \dots, s_u; \sigma|_{\bigcup_{i=1}^u \mathbb{B}(s_i, \tilde{t}_n)}\big)$ for $A \subset \{0,1\}^u$. That is, more exactly, $Q_u^{(n)}\big((s_1, \dots, s_u; \sigma), A\big) = Q_u^{(n)}\big((s_1, \dots, s_u; \sigma'), A\big)$ for $A \subset \{0,1\}^u$ and $(s_1, \dots, s_u; \sigma), (s_1, \dots, s_u; \sigma') \in E_u$ with $\sigma|_{\bigcup_{i=1}^u \mathbb{B}(s_i, \tilde{t}_n)} = \sigma'|_{\bigcup_{i=1}^u \mathbb{B}(s_i, \tilde{t}_n)}$.

Assumption 4 (Control of the short range positive covariances between the retention decisions given $\boldsymbol{\xi}$). For each $n \in \mathbb{N}$, let $\breve{\delta}_n : \mathbb{R}_+ \to \mathbb{R}_+$ be an increasing function such that for every $t \in \mathbb{R}_+$ and for $(s_1, s_2; \sigma) \in E_2$ with $|s_1 - s_2| < \sqrt{D}(t+1)$, we have

$$Q_2^{(n)}\big((s_1, s_2; \sigma), \{(1,1)\}\big) \le \bar{p}_n \breve{\delta}_n(t).$$

Assumption 5 (Control of the long range dependence between the retention decisions given $\boldsymbol{\xi}$). For each $n \in \mathbb{N}$, let $\breve{\gamma}_n : \mathbb{R}_+ \to \mathbb{R}_+$ be a decreasing function such that the following property holds: for every cube $A_{\mathrm{int}} = (a, a + h\mathbf{1})$ with $a \in \mathbb{R}_+^D, h \in (0,1]$, for $t \in \mathbb{R}_+$, for every $(s_1, \dots, s_l, s_{l+1}, \dots, s_u; \sigma) \in E_u$ with $l \ge 1$, $u > l$, $\sum_{i=1}^l \delta_{s_i} = \sigma|_{A_{\mathrm{int}}}$ and $|s_i - s_j| > t$ for $i \in \{1, \dots, l\}$, $j \in \{l+1, \dots, u\}$, and for any set $B \subset \{0,1\}^{\{l+1,\dots,u\}}$, we have

$$\Big| Q_u^{(n)}\Big((s_1, \dots, s_l, s_{l+1}, \dots, s_u; \sigma), \big\{(0, \dots, 0)\big\} \times B\Big) - Q_l^{(n)}\Big((s_1, \dots, s_l; \sigma), \big\{(0, \dots, 0)\big\}\Big)\, Q_{u-l}^{(n)}\Big((s_{l+1}, \dots, s_u; \sigma), B\Big) \Big| \;\le\; l\, \bar{p}_n \breve{\gamma}_n(t).$$

Remark 6.4.C. Assumptions 4 and 5 amount to the statement that, for any $t \in \mathbb{R}_+$ and for any set A_{int} of the above form, given a representation $\sum_{i=1}^\infty \delta_{f_i(\xi)}$ of ξ with measurable functions $f_1, f_2, \dots$ which first enumerate all the points in A_{int} and have images $s_i := f_i(\sigma)$, and given an associated sequence (X_i) of retention decisions with respect to $Q^{(n)}$, we have with $i \ne j$

$$\mathbb{E}(X_i X_j \,|\, \xi = \sigma) \le \bar{p}_n \breve{\delta}_n(t)$$

whenever $|s_i - s_j| < \sqrt{D}(t+1)$, and with $B \subset \{0,1\}^{\{l+1,\dots,u\}}$

$$\Big| \mathrm{cov}\Big(1_{\{\sum_{i=1}^l X_i \ge 1\}}, 1_{\{(X_{l+1}, \dots, X_u) \in B\}} \;\Big|\; \xi = \sigma\Big) \Big| \le\; l\, \bar{p}_n \breve{\gamma}_n(t)$$

whenever $|s_i - s_j| > t$ for all $i \in \{1, \dots, l\}$ and $j \in \{l+1, \dots, u\}$, where $l = \sigma(A_{\mathrm{int}}) \ge 0$.

We are now in the position to formulate the theorem. Again, the corresponding quantitative upper bound can be found at the end of the proof.

Theorem 6.4.D (L_∞-version: upper bound and convergence). *Let ξ be a point process on $\mathbb{R}_+^D$ that satisfies Assumptions 1 and 2″ above, and for each $n \in \mathbb{N}$, let $Q^{(n)} := \big(Q_u^{(n)}\big)$ be an admissible sequence of retention kernels that satisfies Assumptions 3, 4, and 5.*
Set $\underline{\varphi} := \inf_{\boldsymbol{k} \in \mathbb{N}^D} \mathbb{P}[\xi(C_{\boldsymbol{k}}) \ge 1] \in [0,1]$, and let $\tilde{\lambda}_n(\cdot) := \mathbb{E}\big(\int_\cdot Q_1^{(n)}\big((s;\xi), \{1\}\big)\; \xi(ds)\big)$, which is the expectation measure of $\xi_{Q^{(n)}}$. Then, for any $m := m(n) \in \mathbb{N}$ with $m \ge 2\tilde{t}_n$,

we obtain

$$d_2\Big(\mathscr{L}\big(\xi_{Q^{(n)}}\kappa_n^{-1}|_J\big), \mathrm{Po}\big(\tilde{\lambda}_n\kappa_n^{-1}|_J\big)\Big)$$
$$= O\bigg(n^D\bar{p}_n\breve{\delta}_n(0)\,,\; m^D\bigg(n^D \wedge \frac{\log^{\uparrow}(n^D\bar{p}_n)}{\underline{\varphi}\,\underline{p}_n}\bigg)\bar{p}_n\big(\bar{p}_n \vee \breve{\delta}_n(m)\big)\,,$$
$$\sqrt{n^D}\bigg(\sqrt{n^D} \wedge \frac{1}{\sqrt{\underline{\varphi}\,\underline{p}_n}}\bigg)\bar{p}_n\Big(\breve{\beta}^{(\infty)}(m-2\tilde{t}_n) \vee \breve{\gamma}_n(m)\Big)\bigg) \qquad \textit{for } n\to\infty.$$

The right hand side goes to 0 *if, for example,* $\bar{p}_n = O(1/n^D)$, $\breve{\delta}_n(0) = o(1)$, *and there is a sequence* $(m(n))_n$ *with* $m(n) \geq 2\tilde{t}_n$, $m(n) = o\big(1/\bar{p}_n^{1/D}\big)$, *such that* $\breve{\delta}_n(m(n)) = o\big(1/m(n)^D\big)$, $\breve{\gamma}_n(m(n)) = o(1)$, *and* $\breve{\beta}^{(\infty)}(m(n)-2\tilde{t}_n) = o(1)$ *for* $n\to\infty$.

Remark 6.4.E. The main ideas of Theorem 6.4.D and Theorem 6.3.C are closely related. For example, we have formulated in two different ways — once in Assumption 2 and once in Assumptions 2″, 3, and 5 — what is essentially the decreasing dependence between $\big(\xi|_{A_{\mathrm{int}}}, X_1^{(n)}, \ldots, X_{\xi(A_{\mathrm{int}})}^{(n)}\big)$ and $\big(\xi|_{A_{\mathrm{ext}}}, X_{\xi(A_{\mathrm{int}})+1}^{(n)}, \ldots, X_{\xi(A_{\mathrm{int}})+\xi(A_{\mathrm{ext}})}^{(n)}\big)$ with increasing distance between an inner set A_{int} and an outer set A_{ext} and increasing n (where $X_1^{(n)}, \ldots, X_{\xi(A_{\mathrm{int}})}^{(n)}$ and $X_{\xi(A_{\mathrm{int}})+1}^{(n)}, \ldots, X_{\xi(A_{\mathrm{int}})+\xi(A_{\mathrm{ext}})}^{(n)}$ denote the retention decisions for the points in A_{int} and in A_{ext}, respectively).

Nevertheless, there are also substantial differences between the two theorems. We mention briefly two of the more important ones, leaving the confirmation of the formal details for the reader. First, in Theorem 6.3.C the thinning is location-based (i.e. the location in the state space), whereas in Theorem 6.4.D the thinning is point-based (i.e. the point of the original process ξ); that is to say, two or more points of ξ which occupy the same location in the state space must be thinned with the same dependence on other retention decisions and with equal probabilities if we want to apply Theorem 6.3.B, but they can be thinned much more generally if we want to apply Theorem 6.4.D.

Secondly, in Theorem 6.4.D the conditional distribution $\mathscr{L}(X_1, \ldots, X_u \,|\, \xi = \sigma)$ is only allowed to be a function of $\sigma|_{\bigcup_{i=1}^u \mathbb{B}(s_i,\tilde{t}_n)}$, whereas in Theorem 6.3.C it may be a function of all of σ. As an example, consider the π-thinning of a homogenous Poisson process on the real half line with points $S_1 < S_2 < \ldots$, using a retention field of the form

$$\pi_n(\omega, t) := g_1(n)\exp\Big(-g_2(n)\Big/\min_{\substack{i\in\mathbb{N}\\ S_i(\omega)>t}}\big(S_i(\omega)-t\big)\Big)$$

with functions $g_1 : \mathbb{N} \to [0,1]$ and $g_2 : \mathbb{N} \to \mathbb{R}_+$, where the idea is that $g_1(n) \to 0$ and/or $g_2(n) \to \infty$ as $n \to \infty$. Since π_n is only a function of ω through ξ, this situation is easily translated into a Q-thinning model. Obviously, Assumption 3 is not satisfied by this model, because the retention probabilities may depend on arbitrarily distant regions. On the other hand, this long range dependence is very weak, and it can in fact be shown that the conditions for Theorem 6.3.C are met with a mixing coefficient $\breve{\beta}^{(\infty)}(t)$ that does not depend on n and that goes to zero exponentially fast as $t \to \infty$.

Proof of Theorem 6.4.D. Let $\eta_n \sim \mathrm{Po}(\tilde{\lambda}_n)$, and choose an arbitrary $\tilde{n} \in \mathbb{N}$. We use the notation and the conventions from the proofs in Section 6.3, replacing only ξ_{π_n} by $\xi_{Q^{(n)}}$,

and define the concrete "inner" and "outer" sets needed by

$$A_{\mathrm{int}}(\boldsymbol{k},\boldsymbol{r}) := \mathring{C}_{\boldsymbol{kr}}\,,\ A^{(\tilde{t})}_{\mathrm{int}}(\boldsymbol{k}) := \Bigl(\bigcup\nolimits^{n}_{\substack{\boldsymbol{j}=1\\ |\boldsymbol{j}-\boldsymbol{k}|\le\tilde{t}}} C_{\boldsymbol{j}}\Bigr)^{\circ},\ A^{(\tilde{t},t)}_{\mathrm{ext}}(\boldsymbol{k}) := \Bigl(\bigcup\nolimits^{n}_{\substack{\boldsymbol{j}=1\\ |\boldsymbol{j}-\boldsymbol{k}|>\tilde{t}+t}} C_{\boldsymbol{j}}\Bigr)^{\circ}$$

for $\tilde{t}, t \in \mathbb{Z}_+$, $\boldsymbol{k} \in \{1,2,\dots,n\}^D$, and $\boldsymbol{r} \in \{1,2,\dots,\tilde{n}\}^D$. Most of the estimates from the proof of Theorem 6.3.B are still valid for the new thinning $\xi_{Q^{(n)}}$ if we condition on ξ alone instead of ξ and π_n together and replace in the bounds w_1 by $\bar{h}_1\bar{p}_n$ and $\underline{w}_1$ by $\underline{\varphi}\,\underline{p}_n$. Thus, for example,

$$\underline{\varphi}\,\underline{p}_n \le \sum_{r=1}^{\tilde{n}} q_{\boldsymbol{kr}} \le \bar{h}_1\bar{p}_n,$$

$$\begin{aligned} d_2\bigl(\mathscr{L}(\xi_{Q^{(n)}}\kappa_n^{-1}|_J), \mathscr{L}(\Xi_n\kappa_n^{-1})\bigr) + d_2\bigl(\mathscr{L}(\mathrm{H}_n\kappa_n^{-1}), \mathscr{L}(\eta_n\kappa_n^{-1}|_J)\bigr)& \\ \le\ \sqrt{D}/(n\tilde{n}) + \sum_{\boldsymbol{k}=1}^{n}\sum_{\boldsymbol{r}=1}^{\tilde{n}} \mathbb{E}\bigl(\xi_{Q^{(n)}}(C_{\boldsymbol{kr}})1_{\{\xi_{Q^{(n)}}(C_{\boldsymbol{kr}})\ge 2\}}\bigr),& \end{aligned}$$

and also $M_1(\tilde{\mu})$, $M_2(\tilde{\mu})$ and $\sum_{\boldsymbol{k},\boldsymbol{r}} q_{\boldsymbol{kr}}\mathbb{E}\widetilde{Z}_{\boldsymbol{kr}}$ from the local Stein theorem follow exactly this pattern.

The remaining terms can be estimated in a fashion very similar to that of the proof of Theorem 6.3.B, namely

$$\begin{aligned} \sum_{\boldsymbol{r}=1}^{\tilde{n}} \mathbb{E}\bigl(\xi_{Q^{(n)}}(C_{\boldsymbol{kr}})1_{\{\xi_{Q^{(n)}}(C_{\boldsymbol{kr}})\ge 2\}}\bigr) &\le \mathbb{E}\bigl(\xi_{Q^{(n)}}(C_{\boldsymbol{k}})(\xi_{Q^{(n)}}(C_{\boldsymbol{k}})-1)\bigr) \\ &\le \mathbb{E}\Biggl(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})}\sum_{\substack{j=1\\ j\ne i}}^{\xi(C_{\boldsymbol{k}})} \mathbb{E}(X_iX_j\,|\,\xi)\Biggr) \\ &\le \bar{h}_2\bar{p}_n\breve{\delta}_n(0), \end{aligned}$$

and in the local Stein theorem

$$\begin{aligned} \sum_{\boldsymbol{k},\boldsymbol{r}} \mathbb{E}(I_{\boldsymbol{kr}}Z_{\boldsymbol{kr}}) &\le \sum_{\boldsymbol{k}=1}^{n}\Biggl(\sum_{\substack{\boldsymbol{l}=1\\ 1\le|\boldsymbol{l}-\boldsymbol{k}|\le m}}^{n} \mathbb{E}\bigl(\xi_{Q^{(n)}}(C_{\boldsymbol{k}})\,\xi_{Q^{(n)}}(C_{\boldsymbol{l}})\bigr) \ +\ \mathbb{E}\bigl(\xi_{Q^{(n)}}(C_{\boldsymbol{k}})(\xi_{Q^{(n)}}(C_{\boldsymbol{k}})-1)\bigr)\Biggr) \\ &\le \sum_{\substack{\boldsymbol{k},\boldsymbol{l}=1\\ 1\le|\boldsymbol{l}-\boldsymbol{k}|\le m}}^{n} \mathbb{E}\Biggl(\sum_{i=1}^{\xi(C_{\boldsymbol{k}})}\sum_{j=\xi(C_{\boldsymbol{k}})+1}^{\xi(C_{\boldsymbol{k}})+\xi(C_{\boldsymbol{l}})} \mathbb{E}(X_iX_j\,|\,\xi)\Biggr) \ +\ \bar{h}_2n^D\bar{p}_n\breve{\delta}_n(0) \\ &\le \bar{h}_2(2m+1)^Dn^D\bar{p}_n\breve{\delta}_n(m), \end{aligned}$$

using the same partition of the index set Θ as in the proofs in Section 6.3.

Again we have to argue a bit more carefully for the estimation of the $e_{\boldsymbol{kr}}$. If we again set $\mathcal{F}^w_{\boldsymbol{k}} := \mathcal{P}\bigl(\{0,1\}^{\Theta^w_{\boldsymbol{kr}}}\bigr)$ and $W_{\boldsymbol{k}} := (I_{\boldsymbol{ls}})_{(\boldsymbol{l},\boldsymbol{s})\in\Theta^w_{\boldsymbol{kr}}}$, we obtain, in a similar fashion as in the

proof of Theorem 6.3.B,

$$\begin{aligned}
e_{\boldsymbol{kr}}/2 &= \max_{B\in\mathcal{F}^w_{\boldsymbol{k}}}\Big|\mathrm{cov}\big(1_{\{\xi_{Q^{(n)}}(C_{\boldsymbol{kr}})\geq 1\}}\,,\ 1_{\{W_{\boldsymbol{k}}\in B\}}\big)\Big| \\
&\leq \max_{B\in\mathcal{F}^w_{\boldsymbol{k}}}\Big|\mathrm{cov}\Big(\mathbb{P}\big[\xi_{Q^{(n)}}(C_{\boldsymbol{kr}})\geq 1 \bigm| \xi\big]\,,\ \mathbb{P}\big[W_{\boldsymbol{k}}\in B \bigm| \xi\big]\Big)\Big| \\
&\quad + \max_{B\in\mathcal{F}^w_{\boldsymbol{k}}}\Big|\mathbb{E}\Big(\mathrm{cov}\big(1_{\{\xi_{Q^{(n)}}(C_{\boldsymbol{kr}})\geq 1\}}\,,\ 1_{\{W_{\boldsymbol{k}}\in B\}} \bigm| \xi\big)\Big)\Big| \\
&= \max_{B\in\mathcal{F}^w_{\boldsymbol{k}}}\Big|\mathrm{cov}\Big(\mathbb{P}\big[\xi_{Q^{(n)}}(C_{\boldsymbol{kr}})\geq 1 \bigm| \xi|_{A^{(\tilde{t}_n)}_{\mathrm{int}}(\boldsymbol{k})}\big]\,,\ \mathbb{P}\big[W_{\boldsymbol{k}}\in B \bigm| \xi|_{A^{(\tilde{t}_n,m-2\tilde{t}_n)}_{\mathrm{ext}}(\boldsymbol{k})}\big]\Big)\Big| \\
&\quad + \max_{B\in\mathcal{F}^w_{\boldsymbol{k}}}\mathbb{E}\Big|\mathrm{cov}\big(1_{\{\sum_{i=1}^{\xi(C_{\boldsymbol{kr}})}X_i\geq 1\}}\,,\ 1_{\{W_{\boldsymbol{k}}\in B\}} \bigm| \xi\big)\Big|, \qquad (6.15)
\end{aligned}$$

where this time Lemma A.4.A(ii) from Appendix A.4.2 was used. Compared with the earlier proofs, we now have a second term that is in general not zero, because this time the retention decisions need not be independent under the conditioning that we have. In the first summand, we extract $\bar{p}_n|C_{\boldsymbol{kr}}|$ from the first argument of the covariance, and use Assumption 2″; in the second summand we use Assumption 5, and obtain altogether

$$\frac{1}{2}\sum_{\boldsymbol{r}=1}^{\tilde{n}} e_{\boldsymbol{kr}} \leq \bar{p}_n\Big(\breve{\beta}^{(\infty)}(m-2\tilde{t}_n)+\bar{h}_1\breve{\gamma}_n(m)\Big). \qquad (6.16)$$

Once more, assembling the different parts, and letting $\tilde{n}$ go to infinity, gives the overall estimate

$$\begin{aligned}
&d_2\Big(\mathscr{L}\big(\xi_{Q^{(n)}}\kappa_n^{-1}|_J\big),\mathscr{L}\big(\eta_n\kappa_n^{-1}|_J\big)\Big) \\
&\quad\leq \bar{h}_2 n^D\bar{p}_n\breve{\delta}_n(0) \\
&\qquad + \left(1\wedge\left[\frac{2}{\underline{\varphi}n^D\underline{p}_n}\Big(1+2\log^+\Big(\frac{\bar{h}_1}{2}n^D\bar{p}_n\Big)\Big)\right]\right)(2m+1)^D n^D\big(\bar{h}_1^2\bar{p}_n+\bar{h}_2\breve{\delta}_n(m)\big)\bar{p}_n \\
&\qquad + \left(1\wedge\frac{1.65}{(\underline{\varphi}n^D\underline{p}_n)^{1/2}}\right)2n^D\bar{p}_n\Big(\breve{\beta}^{(\infty)}(m-2\tilde{t}_n)+\bar{h}_1\breve{\gamma}_n(m)\Big),
\end{aligned}$$

which is of the required order for $n\to\infty$. □

6.5 Applications

The two examples below are typical applications of the thinning concepts introduced in Sections 6.1 and 6.4. A short description of these applications has already been given in Section 6.2. Very roughly speaking, under the model of Subsection 6.5.1 points are deleted if covered by a certain RACS, which is an example of thinning according to a random environment; whereas under the model of Subsection 6.5.2 points are deleted if they lose a certain kind of competition against other points nearby, which is an example of thinning according to direct point interactions.

6.5.1 Thinning according to a random environment (clouds in the starry sky)

Suppose that the stars within a certain distance $\bar{r}$ to earth, and lying in a large window $J_n \subset \mathbb{R}^2_+$ of the night sky, are part of a point process ξ on $\mathbb{R}^2_+$ that fulfills Assumptions 1 and 2′ from Section 6.3. Whether you can actually see a particular star in this window depends, among other things, upon the distance and the brightness of the star, and upon whether any other object (here, a cloud) covers the star. Suppose, for the sake of simplicity, that the distributions of distance and brightness of stars do not depend on their position in the window, so that there is a basic probability $q_0 = q_0^{(n)}$ that you can see any fixed star in that part of the window that is not covered by clouds. Suppose, furthermore, that the clouds in the sky as seen from earth in the upper right area of some reference point, form a separable RACS $\Xi \subset \mathbb{R}^2_+$ that is independent of ξ. By Remark A.3.B from the appendix, Theorem 6.3.F may then be applied for the retention field given by $\pi_n(\omega, s) = q_0(1 - 1_{\Xi(\omega)}(s))$. In what follows, we admit a general dimension $D \in \mathbb{N}$ for the sky.

In order to make things more concrete, let us consider a toy example. Suppose that the cloud RACS $\Xi = \Xi_n$ is a simple Boolean model consisting only of discs of positive i.i.d. radii whose centers form a homogeneous Poisson process on $\mathbb{R}^D$: set

$$\Xi := \bigcup_{i=1}^{\infty} \big(Y_i + \mathbb{B}(0, R_i^{(n)})\big) \cap \mathbb{R}^D_+, \tag{6.17}$$

where $(Y_i)_{i\in\mathbb{N}}$ are the points of a $\mathrm{Po}(\ell Leb^D)$-process on $\mathbb{R}^D$, $\ell > 0$, and $R_i^{(n)} \in L_\infty$ are i.i.d. $(0,\infty)$-valued random variables that are independent also of the Poisson process with

$$\big\|R_i^{(n)}\big\|_{L_D} =: r_n, \quad \Big(\mathbb{E}\Big(\big(R_i^{(n)} - \sqrt{D}\big)^D 1_{\{R_i^{(n)} > \sqrt{D}\}}\Big)\Big)^{1/D} =: \check{r}_n, \quad \text{and } \big\|R_i^{(n)}\big\|_{L_\infty} =: r_n^{(\infty)}.$$

Note that Ξ is a separable RACS. By Lemma 2.3.E, its capacity functional is given by

$$T_\Xi(C) = 1 - \exp\Big(-\ell\,\mathbb{E}\Big(Leb^D\big(\mathbb{B}(0, R_1^{(n)}) + C\big)\Big)\Big) \tag{6.18}$$

for any compact subset $C \subset \mathbb{R}^D$. Let $\widetilde{\Xi}$ be the RACS that is obtained from Ξ by decreasing by $\sqrt{D}$ the radii $R_i^{(n)}$ of those balls which have $R_i^{(n)} > \sqrt{D}$, and deleting those centers whose balls have $R_i^{(n)} \le \sqrt{D}$. Then, applying (6.18) for $C_{\boldsymbol{k}}$ by continuity of the measure from below (note that $C_{\boldsymbol{k}}$ is not compact),

$$\begin{aligned}
\mathbb{P}\big[\sup_{s\in C_{\boldsymbol{k}}} \pi_n(s) = q_0\big] &= \mathbb{P}\big[C_{\boldsymbol{k}} \not\subset \Xi\big] \le 1 - \mathbb{P}\big[\widetilde{\Xi} \cap C_{\boldsymbol{k}} \neq \emptyset\big] \\
&\le \exp\Big(-\ell\,\mathbb{E}\Big(Leb^D\big(\mathbb{B}(0, R_1^{(n)} - \sqrt{D}) + C_{\boldsymbol{k}}\big) 1_{\{R_i^{(n)} > \sqrt{D}\}}\Big)\Big) \\
&\le \exp\big(-\ell\alpha_D \check{r}_n^D\big),
\end{aligned}$$

where $\alpha_D := \pi^{D/2}/\Gamma(D/2+1)$ is as usual the volume of the D-dimensional unit ball. Therefore

$$\mathbb{E}\Big(\sup_{s\in C_{\boldsymbol{k}}} \pi_n(s)\Big) = q_0 \mathbb{P}\big[\sup_{s\in C_{\boldsymbol{k}}} \pi_n(s) = q_0\big] \le q_0 \exp(-\ell\alpha_D \check{r}_n^D),$$

and

$$\mathbb{E}\Big(\sup_{s\in C_{\boldsymbol{k}}}\pi_n(s)^2\Big) = q_0^2\mathbb{P}\big[\sup_{s\in C_{\boldsymbol{k}}}\pi_n(s) = q_0\big] \le q_0^2\exp(-\ell\alpha_D\check{r}_n^D).$$

Furthermore, we obtain for the approximating expectation measure in Theorem 6.3.F

$$\begin{aligned}\lambda_n\kappa_n^{-1} &= \bigg(\int_{\cdot}\mathbb{E}\pi_n(s)\;\mu_1(ds)\bigg)\kappa_n^{-1} = q_0\exp\big(-\ell\alpha_D r_n^D\big)\mu_1\kappa_n^{-1}\\ &= q_0\exp\big(-\ell\alpha_D r_n^D\big)n^D\int_{\cdot}\big(h_1\circ\kappa_n^{-1}\big)(s)\;Leb^D(ds),\end{aligned}$$

where the second equality is obtained by Equation (2.8), noting that $\mathbb{E}\pi_n(s) = q_0(1-p_\Xi)$ for every $s\in\mathbb{R}_+^D$, where p_Ξ is the volume fraction of $\bigcup_{i=1}^\infty\big(Y_i+\mathbb{B}(0,R_i^{(n)})\big)$. This yields in total the following result.

Proposition 6.5.A. *Let ξ be a point process on $\mathbb{R}_+^D$ that satisfies Assumptions 1 and 2′. Furthermore, let $\pi_n(\omega,s) = q_0(1-1_{\Xi_n(\omega)}(s))$ for all $s\in\mathbb{R}_+^D$, $\omega\in\Omega$, where $q_0 = q_0^{(n)}\in[0,1]$ and Ξ_n is the Boolean model given by Equation (6.17), i.e. the $\mathbb{R}_+^D$-part of a union of balls in $\mathbb{R}^D$ whose centers form a $\mathrm{Po}(\ell Leb^D)$-process and whose radii $R_i^{(n)}$ are positive-valued L_∞ random variables with $r_n^{(\infty)} = \big\|R_i^{(n)}\big\|_{L_\infty}$, $r_n = \big\|R_i^{(n)}\big\|_{L_D}$, and $\check{r}_n = \big(\mathbb{E}\big((R_i^{(n)}-\sqrt{D})^D 1_{\{R_i^{(n)}>\sqrt{D}\}}\big)\big)^{1/D}$.*
Then, for $m = m(n)\in\mathbb{N}$ with $m(n)\ge 2r_n^{(\infty)}$, we obtain that

$$\begin{aligned}&d_2\Big(\mathscr{L}\big(\xi_{\pi_n}\kappa_n^{-1}|_J\big),\mathrm{Po}\big(\lambda_n\kappa_n^{-1}|_J\big)\Big)\\ &\qquad= O\Big(m^D n^D q_0^2\exp\big(-\ell\alpha_D\check{r}_n^D\big)\;,\; n^D q_0\exp\big(-\ell\alpha_D\check{r}_n^D\big)\breve{\beta}^{(\mathrm{ind})}(m)\Big)\quad \text{for } n\to\infty,\end{aligned}$$

where $\lambda_n = q_0\exp\big(-\ell\alpha_D r_n^D\big)\mu_1$ and α_D is the volume of the D-dimensional unit ball.

From an asymptotical point of view, clearly the interesting cases are those in which $\lambda_n\kappa_n^{-1}|_J$ does not fade away to the zero measure as n tends to infinity, giving us an artificial benefit for our distance estimate. In order to prevent this behavior, we must avoid choosing r_n of a higher than logarithmic order. To get a quick idea, consider the special case in which there is a $n_0\in\mathbb{N}$ such that $R_1^{(n)} > \sqrt{D}$ for $n\ge n_0$, and $\mathbb{E}\big((R_1^{(n)})^{D-1}\big) = o\big(\mathbb{E}\big((R_1^{(n)})^D\big)\big)$. This situation allows us to choose an arbitrary $\zeta>0$, and we still find $n_1\in\mathbb{N}$ such that $\check{r}_n^D\ge(1-\zeta)r_n^D$ for all $n\ge n_1$. Let us furthermore arrange for a constant non-zero $\lambda_n\kappa_n^{-1}$: let $\mu_1 := \mu_0 Leb^D$ with $\mu_0>0$, choose $q_0\ge 1/n^D$ and set

$$r_n := \Big(\frac{1}{\ell\alpha_D}\log(q_0 n^D)\Big)^{1/D},$$

so that $\lambda_n\kappa_n^{-1} = \mu_0 Leb^D$ for every $n\in\mathbb{N}$. The result is as follows.

Corollary 6.5.B. *Under the conditions of Proposition 6.5.A, as well as the additional conditions above, we have for any $\zeta>0$*

$$d_2\Big(\mathscr{L}\big(\xi_{\pi_n}\kappa_n^{-1}|_J\big),\mathrm{Po}\big(\mu_0 Leb^D|_J\big)\Big) = O\Big(m^D q_0(n^D q_0)^\zeta\;,\;(n^D q_0)^\zeta\breve{\beta}^{(\mathrm{ind})}(m)\Big)\quad \text{for } n\to\infty.$$

The upper bound in the above proposition goes to zero under an appropriate choice of $m \geq 2r_n^{(\infty)}$ and $\zeta > 0$ if $q_0 = O(n^{-\varepsilon_1 D})$ as $n \to \infty$ and $\breve{\beta}^{(\mathrm{ind})}(t) = O(t^{-\varepsilon_2 D})$ as $t \to \infty$ for some $\varepsilon_1, \varepsilon_2 > 0$. Note that it is always possible to choose m appropriately, provided that $r_n^{(\infty)} = O(n^{\varepsilon_3})$ for some $\varepsilon_3 \leq \varepsilon_1$. In the case of convergence of the above bound to zero we obtain furthermore by Theorem 3.2.A(ii) that

$$\xi_{\pi_n} \kappa_n^{-1}|_J \xrightarrow{\mathscr{D}} \mathrm{Po}\big(\mu_0 Leb^D|_J\big) \quad \text{for } n \to \infty.$$

6.5.2 Thinning according to point interactions (competition in a plant population)

Suppose that the individuals of a certain kind of plant that grow in a large piece $J_n \subset \mathbb{R}^2_+$ of soil are part of a point process $\xi = \sum_{i=1}^{\infty} \delta_{S_i}$ on $\mathbb{R}^2_+$ which has $\sigma(\xi)$-measurable points S_i with realizations s_i, and fulfills Assumptions 1 and $2''$ from Sections 6.3 and 6.4. As before, we will carry out our analysis for a general dimension $D \in \mathbb{N}$. Assume that the plants have certain "fitness parameters" Ψ_i, one per plant, which are i.i.d. $(0,1]$-valued random variables, independent also of everything else and following a continuous distribution function.

Whether a given plant survives until some time t_0 depends first on the overall environmental conditions, which we require to be the same for all plants (say, each plant has a basic survival probability $q_0 = q_0^{(n)}$), and secondly on the influence of other plants in its immediate surroundings. Suppose that the competition is such that an individual plant survives it, independently of whether it survives the environmental effect, if there are no plants with a higher degree of fitness within a radius of $r_n > 0$. We model this situation by assuming that we have for each S_i a retention decision

$$X_i := X_i^{(n)} := Y_i^{(n)} Z_i^{(n)},$$

where the $Y_i^{(n)}$ are i.i.d. $\mathrm{Be}(q_0)$-random variables that are independent of everything else and determine survival due to the environmental effect, and the $Z_i^{(n)}$ are defined by

$$Z_i^{(n)} := \prod_{\substack{j=1 \\ |S_j - S_i| \leq r_n}}^{\infty} 1_{\{\Psi_i \geq \Psi_j\}}$$

and determine survival due to the competition effect. This second thinning effect is the same one used for the construction of the Matern hard core process (see Stoyan, Kendall and Mecke (1987), Section 5.4). We obtain from symmetry considerations

$$\mathbb{E}(X_i \,|\, \xi = \sigma) = q_0 \frac{1}{\sigma\big(\mathbb{B}(s_i, r_n)\big)}, \tag{6.19}$$

and for $i \neq j$,

$$\mathbb{E}(X_i X_j \,|\, \xi = \sigma) = \begin{cases} 0 & \text{if } |s_i - s_j| \leq r_n, \\ \Big(1 + \frac{\sigma(\mathbb{B}(s_i,r_n) \cap \mathbb{B}(s_j,r_n))}{\sigma(\mathbb{B}(s_i,r_n) \cup \mathbb{B}(s_j,r_n))}\Big) q_0^2 \frac{1}{\sigma(\mathbb{B}(s_i,r_n))} \frac{1}{\sigma(\mathbb{B}(s_j,r_n))} & \text{if } |s_i - s_j| > r_n. \end{cases} \tag{6.20}$$

For the second result a bit of computation is necessary, which basically consists of counting the number of orderings of the fitness values of the individuals in $\mathbb{B}(s_i, r_n) \cup \mathbb{B}(s_j, r_n)$ that leave the values of s_i and s_j as highest in their respective "competition ball". Note that, for $|s_i - s_j| > 2r_n$, the retention decisions X_i and X_j are independent given ξ, because the "competition balls" $\mathbb{B}(s_i, r_n)$ and $\mathbb{B}(s_j, r_n)$ are disjoint. By Remark 6.4.A we can define probability kernels by

$$Q_u^{(n)}\big((s_1, \dots, s_u; \sigma), \{(e_1, \dots, e_u)\}\big) := \mathbb{P}[X_1 = e_1, \dots X_u = e_u \,|\, \xi = \sigma]$$

for almost every σ, in such a way that $\big(Q_u^{(n)}\big)_{u\in\mathbb{N}}$ is an admissible family of retention kernels for the sequence $(X_i)_{i\in\mathbb{N}}$.

In view of Theorem 6.4.D, Assumption 3 is satisfied with $\tilde{t}_n = r_n$, and for $m(n) \geq 2r_n$ we can always set $\breve{\gamma}_n(m(n))$ to zero in Assumption 5. Furthermore, we have by Equation (6.19)

$$\bar{p}_n = q_0 \left\| \sup_{s\in E_1(\xi)\cap J_n} \frac{1}{\xi\big(\mathbb{B}(s, r_n)\big)} \right\|_{L_\infty},$$

and obtain, by (6.20) for $\breve{\delta}_n$ in Assumption 4, the possible choice of

$$\breve{\delta}_n :\equiv 2q_0 \left\| \sup_{s\in E_1(\xi)\cap J_n} \frac{1}{\xi\big(\mathbb{B}(s, r_n)\big)} \right\|_{L_\infty}.$$

Obviously, for most point processes ξ, the norms on the right hand side of the above equations are one, and we then do not get very interesting upper bounds for the d_2-distance between the law of the thinned process and a Poisson process law, because the bounds take into account only the environmental thinning effect and not the competition effect.

There are two ways in which this can be rescued. First, we could use the L_1-version of Theorem 6.4.D (i.e. the Q-thinning analog of Corollary 6.3.E), which in this case is much more promising, because we have expectations instead of the L_∞-norms above. For example, instead of $\bar{p}_n$ we would essentially have a term of the form

$$q_0 \sup_{1\leq \boldsymbol{k}\leq n} \mathbb{E}\left(\sup_{s\in C_{\boldsymbol{k}}} \frac{1}{\xi\big(\mathbb{B}(s, r_n)\big)} \right) \leq q_0 \sup_{s\in J_n} \mathbb{E}\left(\frac{1}{\xi\big(\mathbb{B}(s, r_n - \sqrt{D}/2)\big)} \right)$$

(and some correction terms along the lines of Corollary 6.3.E, depending on the number of moments that $\xi(C_{\boldsymbol{k}})$ has). However, as we have not formulated this more involved L_1-version, we do not pursue the idea here any further. Secondly, we can look at a specific ξ-process where we can be sure of having a certain number of competitors in every competition ball: let ξ be the point process on all of $\mathbb{R}^D$ (in order to avoid edge effects) that has in every unit square $[\boldsymbol{k}, \boldsymbol{k} + \mathbf{1})$, $\boldsymbol{k} \in \mathbb{Z}^D$, exactly one point, which is uniformly distributed over the square and independent of the locations of all the other points. We might think of a gardener who sows seeds, exactly one per square, by just throwing each carelessly over its square; the fact that the distribution over each square is uniform is by no means crucial to the essence of the following explanations.

We then have, for all $s \in J_n$,

$$\xi\big(\mathbb{B}(s, r_n)\big) \geq \left(\frac{2r_n}{\sqrt{D}} - 2 \right)^D \geq \left(\frac{r_n}{\sqrt{D}} \right)^D$$

for $r_n \geq 2\sqrt{D}$, and

$$\xi\big(\mathbb{B}(s,r_n)\big) \leq \big(2r_n+2\big)^D \leq (3r_n)^D$$

for $r_n \geq 2$. Therefore, in the notation of Theorem 6.4.D,

$$\bar{p}_n \leq D^{D/2}\frac{q_0}{r_n^D}, \quad \underline{p}_n \geq 3^{-D}\frac{q_0}{r_n^D}, \quad \text{and} \quad \breve{\delta}_n := 2\bar{p}_n \leq 2D^{D/2}\frac{q_0}{r_n^D}.$$

Furthermore, we obviously have

$$\mathbb{E}\xi = Leb^D,$$

and for the second factorial moment measure we can calculate

$$\mu_{[2]}(A\times B) \leq Leb^{2D}(A\times B) \quad \text{for all } A,B\in\mathcal{B}^D,$$

such that $\mu_{[2]}(C)\leq 1$ for every unit cube $C\subset\mathbb{R}^{2D}$. Finally, the mixing property $2''$ is also met with $\breve{\beta}^{(\infty)}(t)=0$ for $t\geq 1$. With $m(n) := \lceil 2r_n\rceil + 1$, this yields the following result.

Proposition 6.5.C. *Let $U_{\boldsymbol{k}}$, $\boldsymbol{k}\in\mathbb{Z}^D$, be independent random variables that are uniformly distributed on $[0,1)^D$, and $\xi := \sum_{\boldsymbol{k}\in\mathbb{Z}^D}\delta_{\boldsymbol{k}+U_{\boldsymbol{k}}}$. Furthermore let $(S_i)_{i\in\mathbb{N}}$ be a $\sigma(\xi)$-measurable enumeration of the points of ξ, and $(\Psi_i)_{i\in\mathbb{N}}$ an i.i.d. sequence of continuous, $(0,1]$-valued random variables that is independent of ξ. Define retention decisions $X_i := X_i^{(n)} := Y_i^{(n)}Z_i^{(n)}$, where $Y_i^{(n)}$, $i\in\mathbb{N}$, are i.i.d. Bernoulli random variables with parameter $q_0 = q_0^{(n)}\in[0,1]$ that are independent of everything else, and $Z_i^{(n)}$, $i\in\mathbb{N}$, are given by*

$$Z_i^{(n)} := \prod_{\substack{j=1\\ |S_j-S_i|\leq r_n}}^{\infty} 1_{\{\Psi_i\geq\Psi_j\}}$$

for some $r_n>0$.
We then have that the point process $\sum_{i=1}^{\infty}X_i\delta_{S_i}$ can be constructed as a Q-thinning $\xi_{Q^{(n)}}$ of ξ for an appropriate sequence $Q=Q^{(n)}$, and obtain

$$d_2\Big(\mathscr{L}\big(\xi_{Q^{(n)}}\kappa_n^{-1}|_J\big), \mathrm{Po}\big(\tilde{\lambda}_n\kappa_n^{-1}|_J\big)\Big) \;=\; O\bigg(n^D\Big(\frac{q_0}{r_n^D}\Big)^2,\; q_0\Big(\Big[n^D\frac{q_0}{r_n^D}\Big]\wedge\log^{\uparrow}\Big(n^D\frac{q_0}{r_n^D}\Big)\Big)\bigg)$$
for $n\to\infty$

with

$$\tilde{\lambda}_n(\cdot) = \mathbb{E}\bigg(\int_{\cdot} Q_1^{(n)}\big((s,\xi),\{1\}\big)\,\xi(ds)\bigg) = q_0\mathbb{E}\bigg(\int_{\cdot}\frac{\xi(ds)}{\xi\big(\mathbb{B}(s,r_n)\big)}\bigg).$$

Note that

$$\tilde{\lambda}_n\kappa_n^{-1}(J) = q_0\mathbb{E}\bigg(\int_{J_n}\frac{\xi(ds)}{\xi\big(\mathbb{B}(s,r_n)\big)}\bigg) \asymp n^D\frac{q_0}{r_n^D} \quad \text{for } n\to\infty,$$

and therefore, in view of obtaining a stable Poisson process, $r_n \asymp nq_0^{1/D}$ is a natural choice, with q_0 not going faster to zero than $1/n^D$. Then

$$d_2\Big(\mathscr{L}\big(\xi_{Q^{(n)}}\kappa_n^{-1}|_J\big), \mathrm{Po}\big(\tilde{\lambda}_n\kappa_n^{-1}|_J\big)\Big) = O\big(q_0\big).$$

Suppose that we would like to compare the distribution of $\xi_{Q^{(n)}\kappa_n^{-1}|_J}$ to the fixed $\mathrm{Po}(Leb^D|_J)$-distribution (any stationary intensity measure $\mu_0 Leb^D$ with $\mu_0 > 0$ can be used as long as r_n is adapted accordingly). This requires a more precise calculation of $\tilde{\lambda}\kappa_n^{-1}$. For any set $A \in \mathcal{B}_J$, we obtain

$$\tilde{\lambda}_n\kappa_n^{-1}(A) = q_0\mathbb{E}\biggl(\int_{nA} \frac{\xi(ds)}{\xi\bigl(\mathbb{B}(s,r_n)\bigr)}\biggr) = q_0\mathbb{E}\biggl(\int_{nA} \frac{\xi(ds)}{\alpha_D r_n^D + \varepsilon_{D,n}(s,\xi)}\biggr),$$

where $\varepsilon_{D,n}(s,\xi)$ is the error we make by just setting $\xi\bigl(\mathbb{B}(s,r_n)\bigr) = Leb^D\bigl(\mathbb{B}(s,r_n)\bigr)$. Note that, for $r_n \geq \sqrt{D}$,

$$\bigl|\varepsilon_{D,n}(s,\xi)\bigr| \leq Leb^D\Bigl(\mathbb{B}\bigl(s,r_n+\sqrt{D}\bigr) \setminus \mathbb{B}(s,r_n)\Bigr) \leq \sqrt{D}\omega_D(2r_n)^{D-1} \leq 3^D\omega_D r_n^{D-1},$$

where $\omega_D = 2\pi^{D/2}/\Gamma(D/2)$ is the surface of the D-dimensional unit ball as usual. Hence

$$\begin{aligned}\frac{|A|}{\alpha_D}n^D\frac{q_0}{r_n^D}\biggl(\frac{\alpha_D r_n^D}{\alpha_D r_n^D + 3^D\omega_D r_n^{D-1}}\biggr) &\leq \tilde{\lambda}_n\kappa_n^{-1}(A)\\ &\leq \frac{|A|}{\alpha_D}n^D\frac{q_0}{r_n^D}\biggl(\frac{\alpha_D r_n^D}{(\alpha_D r_n^D - 3^D\omega_D r_n^{D-1})\vee 0}\biggr).\end{aligned}$$

Thus, if $q_0 = \omega(1/n^D)$ (i.e. $n^D q_0 \to \infty$) for $n \to \infty$, and if we set $r_n := n\bigl((1/\alpha_D)q_0\bigr)^{1/D}$ for all $n \in \mathbb{N}$, we obtain

$$\tilde{\lambda}_n\kappa_n^{-1}(A) \longrightarrow Leb^D(A) \quad \text{for } n \to \infty$$

for all $A \in \mathcal{B}_J$. Hence

$$\mathrm{Po}(\tilde{\lambda}_n\kappa_n^{-1}|_J) \xrightarrow{\mathscr{D}} \mathrm{Po}(Leb^D|_J),$$

and more exactly, by an application of Theorem 4.3.D in combination with Theorem 4.4.D and $d_{BW} \leq d_{TV}$ from Subsection 2.4.4,

$$\begin{aligned}d_2\bigl(\mathrm{Po}(\tilde{\lambda}_n\kappa_n^{-1}|_J), \mathrm{Po}(Leb^D|_J)\bigr) &\leq \bigl|\tilde{\lambda}_n\kappa_n^{-1}(J) - 1\bigr| + \sup_{A\in\mathcal{B}_J}\biggl|\frac{\tilde{\lambda}_n\kappa_n^{-1}(A)}{\tilde{\lambda}_n\kappa_n^{-1}(J)} - Leb^D(A)\biggr|\\ &\leq \frac{3^D\omega_D r_n^{D-1}}{(\alpha_D r_n^D - 3^D\omega_D r_n^{D-1})\vee 0} + \frac{2\cdot 3^D\omega_D r_n^{D-1}}{(\alpha_D r_n^D - 3^D\omega_D r_n^{D-1})\vee 0}\\ &= O\bigl(1/r_n\bigr) \qquad \text{for } n \to \infty.\end{aligned}$$

Thus we have proved the following result.

Corollary 6.5.D. *Under the conditions of Proposition 6.5.C, with $q_0 = \omega(1/n^D)$ for $n \to \infty$ and $r_n := n\bigl((1/\alpha_D)q_0\bigr)^{1/D}$ for every $n \in \mathbb{N}$, we have that*

$$d_2\Bigl(\mathscr{L}\bigl(\xi_{Q^{(n)}\kappa_n^{-1}|_J}\bigr), \mathrm{Po}\bigl(Leb^D|_J\bigr)\Bigr) = O\biggl(\frac{1}{nq_0^{1/D}},\ q_0\biggr) \quad \textit{for } n \to \infty.$$

Chapter 7

Approximating superpostions of point processes

In our final approximation problem, superpositions of possibly dependent point processes on the general compact metric space (E', d_0) are considered. This time using Stein's method for Poisson process approximation directly, we give an estimate for the Barbour-Brown distance between the distribution of such a superposition and an appropriate Poisson process distribution. This estimate is compared to a modern version of Grigelionis' theorem, to a theorem of Banys (1980), and to two distance results in Barbour, Holst, and Janson (1992), including the local Stein Theorem 4.3.A. Furthermore, an application to a spatial birth-death model is presented.

7.1 Introduction to the approximation problem

The superposition of a sequence $(\xi_k)_{k\in\mathbb{N}}$ of point processes was defined in Subsection 2.2.8. In fact, *superposition* is just the historical term for *sum* when the summands are point processes. It is a standard result that the superposition of independent and uniformly sparse processes converges in distribution to a Poisson process as the number of processes and the sparseness increase; a fact which forms for example the theoretical backing for many Poisson models of random occurences in time.

Convergence results of this type were first examined in the context of mass service in telecommunications, with Palm (1943) and Khinchin (1955) being the first sources of formal proofs for Poisson limit theorems, albeit under quite strong assumptions. A general Poisson limit theorem for independent superpositions was then obtained in Grigelionis (1963) for the state space $\mathbb{R}_+$, versions for more universal state spaces in Goldman (1967) and Jagers (1972). We formulate Kallenberg's version of Grigelionis' theorem below. A discussion of results with general infinitely divisible point processes in the limit can be found in Matthes, Kerstan, and Mecke (1978). Note in particular Theorem 3.4.2, which contains Grigelionis' theorem as a special case. All the sources mentioned so far restrict themselves to superpositions of independent point processes. Corresponding results for dependent (mixing) point processes with Poisson and compound Poisson processes in the limit can be found in Banys (1980). A first weak distance estimate for the finite dimensional distributions of an independent superposition on the real line was obtained

in Theorem 2 of Grigelionis (1963). A much stronger result in total variation distance for superpositions of processes with dependent numbers but independent positions of points is an immediate consequence of Theorem 10.H in Barbour, Holst, and Janson (1992). Many rather specialized contributions have to go unmentioned here. For a listing of authors of the more general results left out above, the reader is referred to the historical remarks for Chapter 6 and 7 in Kallenberg (1986) and the introduction of Serfozo (1984) (note the article itself).

In this chapter we give explicit upper bounds for the d_2-distance between the law of a dependent superposition and a Poisson process law, the same setting as in Banys (1980), Section 2. In contrast to the previous two chapters, our main proof idea here is based on a direct application of Stein's method for Poisson process approximation.

A detailed comparison with related results is given in Remark 7.2.D. Our estimate typically performs well compared to these results. It is much more generally applicable than the distance estimates previously obtained (usually in the stronger total variation distance), and it yields bounds that imply convergence under conditions very similar to those of previous limit theorems.

In Section 7.3, we give an application of our upper bound in the context of a spatial birth-death model. We examine the development of an animal population modeled by assigning a birth-death process to each of the individuals. These processes may depend on each other according to the spatial arrangement of the animals. We show that the events occurring in the population over a short period of time are approximately composed of two independent Poisson processes, one for the births and one for the deaths, and give explicit bounds for the approximation.

We finish this section by stating Grigelionis' theorem. First, the useful definition of a null array is given. Let E be a general lcscH.

Definition. For every $n, k \in \mathbb{N}$, let ξ_{nk} be a point process on E. The collection $(\xi_{nk})_{n,k}$ is called a *null array* if

(a) $(\xi_{nk})_{k\in\mathbb{N}}$ is an independent sequence of point processes for every $n \in \mathbb{N}$;

(b) $\sup_{k\in\mathbb{N}} \mathbb{P}[\xi_{nk}(B) \geq 1] \longrightarrow 0$ as $n \to \infty$ for every bounded $B \in \mathcal{B}$.

We now give Grigelionis' theorem in the version of Kallenberg (2002), Theorem 16.18, in order to have some basic possibility of comparison for our result in Section 7.2.

Theorem 7.1.A (Grigelionis). *Let $(\xi_{nk})_{n,k\in\mathbb{N}}$ be a null array of point processes on E. Furthermore, let λ be a locally finite measure on E and denote by η the Poisson process on E with intensity measure λ. Then $\sum_{k=1}^{\infty} \xi_{nk} \xrightarrow{\mathcal{D}} \eta$ for $n \to \infty$ if and only if the following conditions hold:*

(i) $\sum_k \mathbb{P}[\xi_{nk}(B) \geq 1] \longrightarrow \lambda(B)$ $(n \to \infty)$ *for every bounded $B \in \mathcal{B}_\lambda$;*

(ii) $\sum_k \mathbb{P}[\xi_{nk}(B) \geq 2] \longrightarrow 0$ $(n \to \infty)$ *for every bounded $B \in \mathcal{B}$.*

7.2 The main results

We state in this section the main theorem, which gives an upper bound for the distance between the distribution of a superposition and a corresponding Poisson process distribution. Note that this is a static result, so there is no need to have n in our notation, nor is there anything else going to infinity.

Let $(\xi_k)_{k\in\mathbb{N}}$ be a sequence of point processes on the compact metric space (E', d_0) which satisfies $0 < \sum_{k=1}^{\infty} \mathbb{P}[|\xi_k| \geq 1] < \infty$, where $|\xi_k| = \xi_k(E')$. For each $k \in \mathbb{N}$, partition $\mathbb{N}$ as $\big\{\{k\}, \Theta_k^s, \Theta_k^w\big\}$, where the idea is that ξ_l depends strongly on ξ_k for $l \in \Theta_k^s$, and ξ_l depends weakly on ξ_k for $l \in \Theta_k^w$. As usual (see Section 4.3) there is no formal requirement for these partitions: if $\mathbb{N} \setminus \{k\}$ is split up "unnaturally" for many of the k, the bound below is still true, but can be very bad.

For each point process ξ_k, choose a representation as $\xi_k = \sum_{i=1}^{|\xi_k|} \delta_{S_k^{(i)}}$, where $S_k^{(i)}$ are $\sigma(\xi_k)$-measurable random elements for $i \in \mathbb{N}$. That such representations exist follows once more from Corollary 2.2.D. Furthermore set $S_k := S_k^{(1)}$. Then we have the following main result.

Theorem 7.2.A (Principal superposition theorem). *Let $q_k := \mathbb{P}[|\xi_k| \geq 1]$, $q_k' := \mathbb{P}[|\xi_k| \geq 2]$ and define the measure μ on E' by $\mu(B) := \sum_{k=1}^{\infty} \mathbb{P}[|\xi_k| \geq 1, S_k \in B]$ for every $B \in \mathcal{B}$, so that $|\mu| = \sum_{k=1}^{\infty} q_k$. Then*

$$\begin{aligned} &d_2\Big(\mathscr{L}\big(\textstyle\sum_{k=1}^{\infty}\xi_k\big), \mathrm{Po}(\mu)\Big) \\ &\qquad \leq \sum_{k=1}^{\infty} q_k' + M_2(\mu) \sum_{k=1}^{\infty} q_k^2 + M_2(\mu) \sum_{k=1}^{\infty} \sum_{l\in\Theta_k^s} \Big(q_k q_l + \mathbb{P}\big[|\xi_k| \geq 1, |\xi_l| \geq 1\big]\Big) \\ &\qquad\quad + \big(M_1(\mu) + M_2(\mu)\big) \sum_{k=1}^{\infty} \mathbb{E}\Big|\mathbb{P}\big[|\xi_k| \geq 1 \;\big|\; (\xi_l)_{l\in\Theta_k^w}\big] - q_k\Big| \\ &\qquad\quad + M_2(\mu) \sum_{k=1}^{\infty} q_k\, \mathbb{E}\, d_W\Big(\mathscr{L}\big(S_k \;\big|\; |\xi_k| \geq 1\big), \mathscr{L}\big(S_k \;\big|\; |\xi_k| \geq 1, (\xi_l)_{l\in\Theta_k^w}\big)\Big), \end{aligned}$$

where

$$M_1(\mu) = 1 \wedge \frac{1.65}{\sqrt{|\mu|}}, \quad M_2(\mu) = 1 \wedge \left[\frac{2}{|\mu|}\Big(1 + 2\log^+\Big(\frac{|\mu|}{2}\Big)\Big)\right],$$

and d_W denotes the Wasserstein metric on E' with respect to d_0.

Remark 7.2.B (Poisson process with slightly different intensity measure). We have a similar result for $d_2\big(\mathscr{L}(\sum_{k=1}^{\infty}\xi_k), \mathrm{Po}(\tilde{\mu})\big)$ with $\tilde{\mu}(B) = \sum_{k=1}^{\infty} \mathbb{P}[|\xi_k| = 1, \xi_k(B) = 1]$ for every $B \in \mathcal{B}$. Just replace in the above theorem "$|\xi_k| \geq 1$" by "$|\xi_k| = 1$" and μ by $\tilde{\mu}$ every time they occur. The advantage of this alternative result is that no explicit representations of the ξ_k are needed for its formulation.

Corollary 7.2.C. *Let $(\xi_{nk})_{n,k}$ be a null array of point processes and λ a finite measure on E'. In the notation of Theorem 7.2.A, with q_k, q_k' and μ depending now on n, we have*

$$\begin{aligned} &d_2\Big(\mathscr{L}\big(\textstyle\sum_{k=1}^{\infty}\xi_{nk}\big), \mathrm{Po}(\lambda)\Big) \\ &\qquad \leq \sum_{k=1}^{\infty} q_k' + M_2(\mu) \sum_{k=1}^{\infty} q_k^2 \\ &\qquad\quad + \min\left(1, \frac{1.65}{\sqrt{|\mu|}}, \frac{1.65}{\sqrt{|\lambda|}}\right)\Big||\mu| - |\lambda|\Big| + \big(1 - e^{-\min(|\mu|,|\lambda|)}\big)\, d_W\Big(\frac{\mu}{|\mu|}, \frac{\lambda}{|\lambda|}\Big), \end{aligned}$$

which under the Conditions (i) and (ii) of Grigelionis' theorem goes to zero as $n \to \infty$.

Remark 7.2.D (Comparisons with other results).

(a) The sufficiency of Conditions (i) and (ii) in Grigelionis' theorem 7.1.A is by Theorem 3.2.A(ii) an immediate consequence of Corollary 7.2.C.

(b) Theorem 4 of Banys (1980), which like Grigelionis' theorem is a mere convergence result, is not implied directly by Theorem 7.2.A, but the two theorems have very similar flavor. Banys also uses, in indirect form, the concepts of an index set Θ_k^s of strong dependence and an index set Θ_k^w of weak dependence, about which his assumptions are weaker in as far as, for every index k, only $\{1, 2, \ldots, k-1\}$ has to be partitioned, but stronger in as far as there is less freedom in the choice of these partitions. Apart from this difference, however, the summands in our upper bound correspond directly to the terms that have to go to zero in Banys's theorem in order to ensure convergence of the superposition. They are even exactly the same, except for the last two summands, which capture the weak long range dependence: in Banys's theorem this dependence is controlled by the smallness of terms of the form (in our notation)

$$\sum_{k=1}^{\infty} \mathbb{E}\Big|\mathbb{P}\big[\xi_k(B) \geq 1 \,|\, (\xi_l)_{l\in\Theta_k^w}\big] - \mathbb{P}\big[\xi_k(B) \geq 1\big]\Big|$$

for every $B \in \mathcal{B}$.

(c) The setting of Theorem 10.H from Barbour, Holst, and Janson (1992) is a special case of the setting of Theorem 7.2.A above except for the stronger total variation distance that was used there. To cope with this distance, strong assumptions about the independence of the point positions were made in Theorem 10.H, which we do not need for our Theorem 7.2.A (note that the motivation for Theorem 10.H was a very different one). The basic ideas of the proofs are the same in both theorems. Under the more restrictive setting of Theorem 10.H, the upper bounds obtained for the two distances are also the same, up to some rather slight differences in the factors $M_1(\mu)$ and $M_2(\mu)$.

(d) Another special case, which can be compared more directly, is the local Stein Theorem 4.3.A. In that theorem, an upper bound is given for the d_2-distance between a dependent indicator point process $\sum_{k'\in\Theta'} I_{k'}\delta_{\alpha_{k'}}$ and a Poisson process $\sum_{k'\in\Theta'} Y_{k'}\delta_{\alpha_{k'}}$ with the same expectation measure λ, where $\{\alpha_{k'}; k' \in \Theta'\} \subset E'$ for a finite nonempty index set Θ'. We enumerate the elements of Θ' by $r(1), \ldots, r(j)$, and define point processes $\xi_k := I_{r(k)}\delta_{\alpha_{r(k)}}$ for $k \leq j$ and $\xi_k := 0$ for $k > j$, where $k \in \mathbb{N}$. We thus obtain the indicator process $\sum_{k'\in\Theta'} I_{k'}\delta_{\alpha_{k'}}$ as a very special case of our superposition $\sum_{k=1}^{\infty} \xi_k$ (there is at most one point per point process, and its position is deterministic). Additionally, we have $\lambda = \mu$ for the corresponding measure μ from Theorem 7.2.A, so that $\sum_{k'\in\Theta'} Y_{k'}\delta_{\alpha_{k'}} \sim \mathrm{Po}(\mu)$. Consequently, Theorem 7.2.A can be applied for comparing the distributions of the indicator process and the Poisson process, yielding an upper bound that is qualitatively exactly the same as the one from Chapter 4 and differs in absolute terms only by having the factor $\big(M_1(\mu) + M_2(\mu)\big)$ instead of only $M_1(\mu)$ in front of the fourth summand. So, up to this changed factor, Theorem 4.3.A is contained in Theorem 7.2.A as a special case. The latter result is strictly more general, among other things in that the

neighborhoods of strong dependence Θ_k^s are not bound to fixed regions of the state space.

Proof of Theorem 7.2.A

Our strategy is to reduce the point processes ξ_k to their first points, and then apply Stein's method to the superposition of these reduced point processes. Let η be a Poisson point process with intensity measure μ, and split up the initial distance as

$$d_2\Big(\mathscr{L}\big(\textstyle\sum_k \xi_k\big), \mathscr{L}(\eta)\Big) \;\le\; d_2\Big(\mathscr{L}\big(\textstyle\sum_k \xi_k\big), \mathscr{L}\big(\textstyle\sum_k I_k\delta_{S_k}\big)\Big) + d_2\Big(\mathscr{L}\big(\textstyle\sum_k I_k\delta_{S_k}\big), \mathscr{L}(\eta)\Big),$$

where $I_k := 1_{\{|\xi_k|\ge 1\}}$ for every $k \in \mathbb{N}$.

The reduction term is very easily estimated, using the "natural coupling" of the two processes. By Theorem 3.2.A(i) we have

$$\begin{aligned}
&d_2\Big(\mathscr{L}\big(\textstyle\sum_k \xi_k\big), \mathscr{L}\big(\textstyle\sum_k I_k\delta_{S_k}\big)\Big) \\
&\qquad\le\; \mathbb{E}\, d_1\big(\textstyle\sum_k \xi_k, \sum_k I_k\delta_{S_k}\big) \\
&\qquad=\; \mathbb{P}\big[|\textstyle\sum_k \xi_k| \ne |\sum_k I_k\delta_{S_k}|\big] + \mathbb{E}\Big(d_1\big(\textstyle\sum_k \xi_k, \sum_k I_k\delta_{S_k}\big)\, 1_{\{|\sum_k \xi_k| = |\sum_k I_k\delta_{S_k}|\}}\Big) \\
&\qquad=\; \mathbb{P}\big[\textstyle\bigcup_k \{|\xi_k| \ge 2\}\big] \\
&\qquad\le\; \sum_k q_k', \qquad (7.1)
\end{aligned}$$

where the expectation in the third line is zero, because $I_k\delta_{S_k} \le \xi_k$ for every $k \in \mathbb{N}$.

For the distance between the distributions of the reduced superposition and the Poisson process we apply Stein's method for Poisson process approximation, following the general procedure presented in Section 4.2. Set $\Xi := \sum_{k=1}^\infty I_k\delta_{S_k}$, and $\Xi_k := \sum_{l\in\mathbb{N}\setminus\{k\}} I_l\delta_{S_l}$ and $\Xi_k^w := \sum_{l\in\Theta_k^w} I_l\delta_{S_l}$ for every $k \in \mathbb{N}$. Choose random elements $\widetilde{S}_1, \widetilde{S}_2, \ldots$ in E' that are independent among each other and of anything else, such that $\widetilde{S}_k \sim \mathscr{L}(S_k \,|\, I_k = 1)$. Fix $f \in \mathfrak{F}_2$ and let $h = h_f$ be the solution to the Stein equation (4.5) given by (4.6). We then have

$$\begin{aligned}
&\big|\mathbb{E}f(\Xi) - \mathbb{E}f(\eta)\big| \\
&\qquad= \Big|\mathbb{E}\int_{E'} \big[h(\Xi - \delta_s) - h(\Xi)\big]\, \Xi(ds) + \mathbb{E}\int_{E'} \big[h(\Xi + \delta_s) - h(\Xi)\big]\, \mu(ds)\Big| \\
&\qquad= \Big|\mathbb{E}\Big(\sum_{k=1}^\infty I_k\big[h(\Xi - \delta_{S_k}) - h(\Xi)\big]\Big) + \mathbb{E}\Big(\sum_{k=1}^\infty q_k\, \mathbb{E}\big(h(\Xi + \delta_{\widetilde{S}_k}) - h(\Xi) \,\big|\, \Xi\big)\Big)\Big| \\
&\qquad\le \sum_{k=1}^\infty \Big|\mathbb{E}\Big(I_k\big[h(\Xi - \delta_{S_k}) - h(\Xi)\big]\Big) + \mathbb{E}\Big(q_k\big[h(\Xi + \delta_{\widetilde{S}_k}) - h(\Xi)\big]\Big)\Big|,
\end{aligned}$$

where we used that $\int g(x)\, \mu(dx) = \sum_{k=1}^\infty \int g(x)\, \mu_k(dx)$ for $\mu_1, \mu_2, \ldots$ and $\mu = \sum_{k=1}^\infty \mu_k$ finite measures on E' and $g \in L_1(\mu)$ in the third line, and Fubini's theorem in the last

line (both based on $\sum_k \mathbb{P}[|\xi_k| \geq 1] < \infty$ and $\|\Delta_1 h\|_\infty \leq 1$). The k-th summand can then be split up further as

$$\begin{aligned}
&\Big|\mathbb{E}\Big(I_k\big[h(\Xi - \delta_{S_k}) - h(\Xi)\big]\Big) - \mathbb{E}\Big(q_k\big[h(\Xi) - h(\Xi + \delta_{\widetilde{S}_k})\big]\Big)\Big| \\
&\qquad \leq \Big|\mathbb{E}\Big(I_k\big[h(\Xi - \delta_{S_k}) - h(\Xi)\big]\Big) - \mathbb{E}\Big(I_k\big[h(\Xi_k^w) - h(\Xi_k^w + \delta_{S_k})\big]\Big)\Big| \\
&\qquad\quad + \Big|\mathbb{E}\Big(I_k\big[h(\Xi_k^w) - h(\Xi_k^w + \delta_{S_k})\big]\Big) - \mathbb{E}\Big(I_k\big[h(\Xi_k^w) - h(\Xi_k^w + \delta_{\widetilde{S}_k})\big]\Big)\Big| \\
&\qquad\quad + \Big|\mathbb{E}\Big(I_k\big[h(\Xi_k^w) - h(\Xi_k^w + \delta_{\widetilde{S}_k})\big]\Big) - \mathbb{E}\Big(q_k\big[h(\Xi_k^w) - h(\Xi_k^w + \delta_{\widetilde{S}_k})\big]\Big)\Big| \\
&\qquad\quad + \Big|\mathbb{E}\Big(q_k\big[h(\Xi_k^w) - h(\Xi_k^w + \delta_{\widetilde{S}_k})\big]\Big) - \mathbb{E}\Big(q_k\big[h(\Xi) - h(\Xi + \delta_{\widetilde{S}_k})\big]\Big)\Big|. \qquad (7.2)
\end{aligned}$$

The first summand in Inequality (7.2)

Assume without loss of generality that Θ_k^s is an infinite set, for if it is not, we can add infinitely many 0-processes to the superposition and put all their indices into Θ_k^s. Enumerate the elements in Θ_k^s by $r(1), r(2), \ldots$ and write $\Xi_k^{w,j} := \Xi_k^w + \sum_{i=1}^j I_{r(i)}\delta_{S_{r(i)}}$ for $j \geq 0$. Since by Borel-Cantelli $\sum_{i=1}^\infty I_{r(i)}$ is almost surely finite, the first summand can be expanded into a telescopic sum, and is hence equal to

$$\begin{aligned}
&\Big|\mathbb{E}\Big(I_k \sum_{j=1}^\infty \Big(\big[h(\Xi_k^{w,j}) - h(\Xi_k^{w,j} + \delta_{S_k})\big] - \big[h(\Xi_k^{w,j-1}) - h(\Xi_k^{w,j-1} + \delta_{S_k})\big]\Big)\Big)\Big| \\
&\qquad\qquad \leq \sum_{j=1}^\infty \Big|\mathbb{E}\Big(I_k\big[h(\Xi_k^{w,j}) - h(\Xi_k^{w,j} + \delta_{S_k}) - h(\Xi_k^{w,j-1}) + h(\Xi_k^{w,j-1} + \delta_{S_k})\big]\Big)\Big|.
\end{aligned}$$

The moduli can be further bounded by

$$\mathbb{E}\Big(I_k I_{r(j)}\Big|h(\Xi_k^{w,j-1} + \delta_{S_{r(j)}} + \delta_{S_k}) - h(\Xi_k^{w,j-1} + \delta_{S_{r(j)}}) - h(\Xi_k^{w,j-1} + \delta_{S_k}) + h(\Xi_k^{w,j-1})\Big|\Big),$$

such that by Inequality (4.8) the total bound for the first summand in Inequality (7.2) is

$$M_2(\mu)\sum_{j=1}^\infty \mathbb{E}(I_k I_{r(j)}) = M_2(\mu)\sum_{l \in \Theta_k^s} \mathbb{E}(I_k I_l).$$

The second summand in Inequality (7.2)

We first show that for any $\varrho \in \mathfrak{N}'$ the function $g_\varrho : E' \to \mathbb{R}$ given by

$$g_\varrho(s) := h(\varrho + \delta_s) \quad \text{for all } s \in E'$$

is d_0-Lipschitz continuous with constant $C := 1 \wedge \frac{1}{|\mu|}\big(\log^+(|\mu|) + 1\big)$. This is done in a way similar to the derivation of the bounds (4.7) and (4.8). Write the spatial immigration-death processes Z and Z' with (deterministic) initial configurations $\varrho + \delta_s$ and $\varrho + \delta_{s'}$ as $Z_1 + \delta_s 1_{\{E>t\}}$ and $Z_1 + \delta_{s'} 1_{\{E>t\}}$, respectively, where E is a standard exponentially distributed random variable that is independent of everything else, and Z_1 is the immigration-death process with immigration measure μ and unit per-capita death rate that starts with

configuration ϱ. We furthermore write Z_0 for the same immigration-death process that starts with zero points. Note that $Z_0(t) \sim \mathrm{Po}((1-e^{-t})\mu)$, and write $\mu_t := (1-e^{-t})|\mu|$. Then, using the explicit form of h given by Equation (4.6), we have that

$$\begin{aligned}
\big|g_\varrho(s) - g_\varrho(s')\big| &= \left|\int_0^\infty \Big(\mathbb{E}f\big(Z(t)\big) - \mathbb{E}f\big(Z'(t)\big)\Big)\, dt\right| \\
&= \left|\int_0^\infty \Big(\mathbb{E}f\big(Z_1(t)+\delta_s\big) - \mathbb{E}f\big(Z_1(t)+\delta_{s'}\big)\Big)\, \mathbb{P}[E>t]\, dt\right| \\
&\le \int_0^\infty \mathbb{E}\, d_1\big(Z_1(t)+\delta_s, Z_1(t)+\delta_{s'}\big)\, e^{-t}\, dt \\
&= d_0(s,s') \int_0^\infty \mathbb{E}\bigg(\frac{1}{|Z_1(t)|+1}\bigg) e^{-t}\, dt \qquad (7.3)
\end{aligned}$$

by Lemma 3.1.A(ii), where furthermore

$$\mathbb{E}\bigg(\frac{1}{|Z_1(t)|+1}\bigg) \le \mathbb{E}\bigg(\frac{1}{|Z_0(t)|+1}\bigg) = \frac{1-e^{-\mu_t}}{\mu_t}.$$

Hence it follows that the integral at the end of Inequality (7.3) is bounded by C, which yields the required Lipschitz continuity.

The second term in Inequality (7.2) is now estimated for $q_k > 0$ as

$$\begin{aligned}
&\Big|\mathbb{E}\Big(I_k\big[h(\Xi_k^w) - h(\Xi_k^w + \delta_{S_k})\big]\Big) - \mathbb{E}\Big(I_k\big[h(\Xi_k^w) - h(\Xi_k^w + \delta_{\widetilde{S}_k})\big]\Big)\Big| \\
&\qquad = \Big|\mathbb{E}\Big(I_k\big[g_{\Xi_k^w}(\widetilde{S}_k) - g_{\Xi_k^w}(S_k)\big]\Big)\Big| \\
&\qquad = \Big|\mathbb{E}\Big(I_k\, \mathbb{E}\Big(g_{\Xi_k^w}(\widetilde{S}_k) - g_{\Xi_k^w}(S_k) \;\Big|\; I_k = 1, \Xi_k^w\Big)\Big)\Big| \\
&\qquad \le C\, \mathbb{E}\Big(I_k\, d_W\Big(\mathscr{L}(S_k \,|\, I_k = 1), \mathscr{L}(S_k \,|\, I_k = 1, \Xi_k^w)\Big)\Big) \\
&\qquad \le C\, q_k\, \mathbb{E}\, d_W\Big(\mathscr{L}(S_k \,|\, I_k = 1), \mathscr{L}(S_k \,|\, I_k = 1, \Xi_k^w)\Big) + C\, \mathbb{E}\big|\mathbb{E}(I_k \,|\, \Xi_k^w) - q_k\big|,
\end{aligned}$$

where for the third line we used that $I_k\, F(I_k, X) = I_k\, F(1, X)$ for any random variable X and any suitable function F. Note that the d_W-term is a measurable function in Ξ_k^w, because the supremum in its definition can be substituted by the supremum over a countable set of functions. The overall bound above is trivially true for $q_k = 0$ as well.

The third summand in Inequality (7.2)

Since $\widetilde{S}_k$ is independent of (I_k, Ξ_k^w) and hence $\widetilde{S}_k \perp\!\!\!\perp_{\Xi_k^w} I_k$ (i.e. $\widetilde{S}_k$ is independent of I_k given Ξ_k^w), we obtain for the third summand

$$\begin{aligned}
&\Big|\mathbb{E}\Big(I_k\big[h(\Xi_k^w) - h(\Xi_k^w + \delta_{\widetilde{S}_k})\big]\Big) - \mathbb{E}\Big(q_k\big[h(\Xi_k^w) - h(\Xi_k^w + \delta_{\widetilde{S}_k})\big]\Big)\Big| \\
&\qquad = \Big|\mathbb{E}\Big(\mathbb{E}\Big((I_k - q_k)\big[h(\Xi_k^w) - h(\Xi_k^w + \delta_{\widetilde{S}_k})\big] \;\Big|\; \Xi_k^w\Big)\Big)\Big| \\
&\qquad = \Big|\mathbb{E}\Big(\mathbb{E}\big(I_k - q_k \;\big|\; \Xi_k^w\big)\, \mathbb{E}\big(h(\Xi_k^w) - h(\Xi_k^w + \delta_{\widetilde{S}_k}) \;\big|\; \Xi_k^w\big)\Big)\Big| \\
&\qquad \le M_1(\mu)\, \mathbb{E}\big|\mathbb{E}(I_k \,|\, \Xi_k^w) - q_k\big|
\end{aligned}$$

by Inequality (4.7).

The fourth summand in Inequality (7.2)

We proceed in the same way as for the first summand and use the corresponding notation. Thus the fourth summand can be expanded into a telescopic sum and hence estimated by

$$q_k \sum_{j=1}^{\infty} \Big| \mathbb{E}\Big(\big[h(\Xi_k^{w,j}) - h(\Xi_k^{w,j} + \delta_{\widetilde{S}_k})\big] - \big[h(\Xi_k^{w,j-1}) - h(\Xi_k^{w,j-1} + \delta_{\widetilde{S}_k})\big] \Big) \Big| \\ + q_k \Big| \mathbb{E}\Big(\big[h(\Xi) - h(\Xi + \delta_{\widetilde{S}_k})\big] - \big[h(\Xi_k) - h(\Xi_k + \delta_{\widetilde{S}_k})\big] \Big) \Big|.$$

We bound the moduli in the analogous way as for the first summand and have thus

$$M_2(\mu) \Big(\sum_{l \in \Theta_k^s} q_k q_l + q_k^2 \Big)$$

as the total bound for the fourth summand.

Assembling of the four estimates for the summands in Inequality (7.2) yields

$$\begin{aligned} d_2\Big(\mathscr{L}\big(\textstyle\sum_k I_k \delta_{S_k}\big), \mathscr{L}(\eta) \Big) &= \sup_{f \in \mathfrak{F}_2} \big| \mathbb{E} f(\Xi) - \mathbb{E} f(\eta) \big| \\ &\le M_2(\mu) \sum_{k=1}^{\infty} q_k^2 + M_2(\mu) \sum_{k=1}^{\infty} \sum_{l \in \Theta_k^s} \big(q_k q_l + \mathbb{E}(I_k I_l) \big) \\ &\quad + (M_1(\mu) + C) \sum_{k=1}^{\infty} \mathbb{E} \big| \mathbb{E}(I_k \,|\, \Xi_k^w) - q_k \big| \\ &\quad + C \sum_{k=1}^{\infty} q_k \, \mathbb{E}\, d_W\Big(\mathscr{L}(S_k \,|\, I_k = 1), \mathscr{L}(S_k \,|\, I_k = 1, \Xi_k^w) \Big). \end{aligned}$$

Together with Inequality (7.1), and noting that $C \le M_2(\mu)$ and that Ξ_k^w is a measurable function of $(\xi_l)_{l \in \Theta_k^w}$, we obtain the required result. □

Proof of Corollary 7.2.C

Choose $\Theta_k^s := \emptyset$ and $\Theta_k^w := \Theta_k = \mathbb{N} \setminus \{k\}$ in Theorem 7.2.A, so that the last three terms in the upper bound of the theorem disappear. We are left with the first two terms and the additional term of $d_2\big(\mathrm{Po}(\mu), \mathrm{Po}(\lambda)\big)$, which can be estimated by Theorem 4.3.D. Thus, we obtain the required bound.

For the convergence statement note that Condition (i) in Theorem 7.1.A implies for $|\mu|, |\lambda| > 0$ that $\frac{\mu}{|\mu|} = \frac{\mu_n}{|\mu_n|}$ converges weakly to $\frac{\lambda}{|\lambda|}$ for $n \to \infty$. Hence the fourth summand in the upper bound goes to zero by Theorem 2.4.D, taking into account that d_W is also the bounded Wasserstein metric because of $d_0 \le 1$. The first and third summands go to zero directly by Conditions (ii) and (i), respectively. Finally, the second summand can be estimated as $M_2(\mu)|\mu| \sup_{k \ge 1} q_k$, which goes to zero, because $(\xi_{nk})_{n,k}$ is a null array, and $|\mu|$ goes to $|\lambda|$. □

7.3 Application: short term behavior of a spatial birth-death model

Consider a population of n animals that are more or less evenly spread over a certain area. The k-th animal and its offspring at time t are described by a birth-death process (BDP) ζ_k with starting state 1 (see e.g. Feller (1968), Section XVII.5 for the definition). The corresponding processes may be rather strongly dependent if two animals live close together, but the dependence between processes is expected to decay with the distance between animals.

In what follows, we provide a rather general modeling framework for this situation and derive statements about the short-term behavior of populations of the above form. We use biological terminology (such as "animals" or "predators") for illustrative purposes, but it should be noted that the model we present is still somewhat too abstract for a serious modeling attempt of any concrete biological situation. On the other hand, the model is flexible enough to be adapted to many other contexts where a birth-death paradigm is reasonable. Examples include the failure and repair of components in complex systems, attacks in computer networks, absorption and emission of photons or other particles, or arrivals and departures in a large system of queues.

Our concrete model is as follows. Let there be an infinite number of animals in $\mathbb{R}_+^D$ represented by a point measure $\varrho = \sum_{k=1}^{\infty} \delta_{z_k}$. The population described above will consist of the first n of these points, so it is preferable to number them in a reasonable way, for example according to their distance from the origin. We assume that this infinite group of animals is "evenly spread", meaning that there is a constant $\overline{c} > 0$ such that

$$\varrho(\mathbb{B}(z_k, r)) - 1 \leq \overline{c} r^D \tag{7.4}$$

for all $r \geq 0$ and $k \in \mathbb{N}$, where $\mathbb{B}(z, r)$ denotes the closed Euclidean ball with center in z and radius r. Depending on the situation, other metrics on the set $\{z_k; k \in \mathbb{N}\}$ and more general functions in r bounding the left hand side of (7.4) might be more appropriate.

Let ζ_k, $k \in \mathbb{N}$, be identically distributed BDPs with birth rates $(\alpha_j)_{j \in \mathbb{Z}_+}$ and death rates $(\beta_j)_{j \in \mathbb{Z}_+}$ which all start with one individual at time 0. We think of ζ_k as the process that belongs to the original animal at z_k. The dependence between these processes is controlled by functions $\phi : [0,1] \to [0,1]$ and $\psi_1, \psi_2 : [0,1] \times \mathbb{R}_+ \to [0,1]$ which are chosen in such a way that

$$\mathbb{P}\big[\zeta_k \text{ and } \zeta_l \text{ each have a jump in } [0,t]\big] \leq t\phi(t) \tag{7.5}$$

for $k \neq l$, and

$$\sup_{F \in \mathcal{F}_k(t)} \mathbb{E}\Big|\mathbb{P}[F \,|\, \mathcal{G}_k(t,r)] - \mathbb{P}[F]\Big| \leq \psi_1(t,r) \tag{7.6}$$

$$\text{and} \quad \mathbb{E}\bigg(\operatorname*{ess\,sup}_{F \in \mathcal{F}_k(t)} \Big|\mathbb{P}[F \,|\, \mathcal{G}_k(t,r)] - \mathbb{P}[F]\Big|\bigg) \leq \psi_2(t,r) \tag{7.7}$$

for every k, where $\mathcal{F}_k(t) := \sigma(\zeta_k|_{[0,t]})$ and $\mathcal{G}_k(t,r) := \sigma(\zeta_l|_{[0,t]};\ l \in \mathbb{N}, |z_l - z_k| > r)$. For the definition of the essential supremum of an arbitrary set of random variables see Neveu (1965), Proposition II.4.1.

The idea behind this dependence structure is as follows: ϕ controls the short-term positive correlation of events (births or deaths) happening at points close together. In the

biological setting, deaths of animals living close together, for example, might be rather strongly positively correlated, because they might be caused (among other reasons) by predators roaming the neighborhood or by fights among the animals. On the other hand, the functions ψ_1 and ψ_2 control the short-term dependence over long distances by providing bounds for the α- and β-mixing coefficients, respectively, between the evolution of a single process and the evolution of all the processes far enough away. See Doukhan (1994), Section 1.1 for an introduction to mixing coefficients. Note that the term bounded by ψ_1 in Inequality (7.6) is in fact twice the α-mixing condition used by Doukhan and elsewhere. In the animal framework, the long-range dependence might be caused by the abundance or scarcity of prey or by the environmental conditions, such as climate or vegetation.

Consider now the population of the first n animals. Theorem 7.2.A yields a result concerning the aggregated population process, observed over a short period of time $h > 0$. Denote by $\widetilde{\zeta}_k$ the event point process for the BDP ζ_k, which we define as the point process on $\mathbb{R}_+ \times \{0,1\}$ that has a point in (t,u) (in other words: a point in t with mark u) if ζ_k has an event of type u at time t, where $u = 1$ codes for a birth and $u = 0$ codes for a death. We choose the distance d_0 on $E' := [0,1] \times \{0,1\}$ that is defined by $d_0\big((t_1,u_1),(t_2,u_2)\big) := \max\big(|t_2 - t_1|, |u_2 - u_1|\big)$ for all $(t_1,u_1),(t_2,u_2) \in E'$.

Proposition 7.3.A. *Suppose that the conditions above hold, that is, let $(\zeta_k)_{k\in\mathbb{N}}$ be identically distributed birth-death processes, attached to the points z_k of a point measure ϱ that satisfies Inequality (7.4) for some $\overline{c} > 0$, and with birth and death rates $(\alpha_j)_{j\in\mathbb{Z}_+}$ and $(\beta_j)_{j\in\mathbb{Z}_+}$, respectively, and let the dependence between the ζ_k be controlled by Inequalities (7.5), (7.6), and (7.7). Define furthermore, for $n \in \mathbb{N}$ and $h > 0$, the Poisson intensity measure $\lambda_{n,h}$ on E' by $\lambda_{n,h} := nh\big(\beta_1(Leb \otimes \delta_0) + \alpha_1(Leb \otimes \delta_1)\big)$ and the time dilation function $\chi_h : \mathbb{R}_+ \times \{0,1\} \to \mathbb{R}_+ \times \{0,1\}$ by $\chi_h(t,u) := (t/h, u)$. Write $\xi_k^{(h)} := \widetilde{\zeta}_k \chi_h^{-1}|_{E'}$ for the dilated point process of the events of ζ_k up to time h. Then there is a constant $K > 0$, such that*

$$d_2\Big(\mathscr{L}\big(\textstyle\sum_{k=1}^n \xi_k^{(h)}\big), \mathrm{Po}(\lambda_{n,h})\Big) \\ \leq K \inf_{r\geq 0}\left(nh^2 + \log^{\uparrow}(nh) r^D \big(h \vee \phi(h)\big) + \sqrt{nh}\frac{\psi_1(h,r)}{h} + \log^{\uparrow}(nh)\frac{\psi_2(h,r)}{h}\right)$$

for any $n \in \mathbb{N}$ and any $h \in \big(0, \frac{1}{\alpha_1+\beta_1}\big)$, where $\log^{\uparrow}(x) := 1 + \log^+(x)$ for $x > 0$.

An upper bound with explicit constants for general $h > 0$, which furthermore improves considerably on the above bound for small nh, can be found at the end of the proof, in Inequality (7.14). To make the result more transparent, we consider the special case where $h = h_n = 1/n$ and some of the other conditions are simplified as well.

Corollary 7.3.B. *Suppose that the conditions of Proposition 7.3.A hold, that additionally $h = h_n = 1/n$ for all $n \in \mathbb{N}$, and that there is a function $\psi : \mathbb{R}_+ \to [0,1]$ with $\psi_2(t,r) \leq t\psi(r)$ for $t \in [0,1]$, $r \in \mathbb{R}_+$. Note that $\lambda_{n,1/n} =: \lambda$ does not depend on n now. Then there is a constant $K > 0$, such that*

$$d_2\Big(\mathscr{L}\big(\textstyle\sum_{k=1}^n \xi_k^{(1/n)}\big), \mathrm{Po}(\lambda)\Big) \leq K \inf_{r\geq 0}\left(\frac{r^D + 1}{n} + r^D\phi(1/n) + \psi(r)\right)$$

for any $n \in \mathbb{N}$. For $\phi(t) = O(t^a)$ $(t \to 0)$ and $\psi(r) = O(r^{-bD})$ $(r \to \infty)$ with constants $a, b > 0$, we have

$$d_2\Big(\mathscr{L}\big(\textstyle\sum_{k=1}^n \xi_k^{(1/n)}\big), \mathrm{Po}(\lambda)\Big) = O\big(n^{-(a\wedge 1)b/(1+b)}\big) \quad \textit{for } n \to \infty,$$

and by Theorem 3.2.A(ii) that

$$\sum_{k=1}^n \xi_k^{(1/n)} \xrightarrow{\mathscr{D}} \mathrm{Po}(\lambda) \quad \textit{for } n \to \infty.$$

□

Proof of Proposition 7.3.A. A few adaptations are necessary for the application of Theorem 7.2.A. Fix $n \in \mathbb{N}$ and $h > 0$. Write $\sigma_j := \alpha_j + \beta_j$ for all $j \in \mathbb{Z}_+$. We exclude a trivial case by assuming that $\sigma_1 > 0$. Write furthermore T_k for the time of the first event of ζ_k, T_k' for the time between the first and the second event of ζ_k, and U_k and U_k' for the types (0 or 1) of these events, respectively. Note that T_k' might be infinite if σ_0 or σ_2 is zero. Set $S_k := \chi_h\big(T_k \wedge h, U_k\big)$ and $S_k' := \chi_h\big((T_k + T_k') \wedge h, U_k'\big)$, and write $\xi_k := \xi_k^{(h)}$ and $\lambda := \lambda_{n,h}$. In order to obtain the right Poisson intensity measure for the theorem, we split up the initial distance as

$$d_2\Big(\mathscr{L}\big(\textstyle\sum_{k=1}^n \xi_k\big), \mathrm{Po}(\lambda)\Big) \le d_2\Big(\mathscr{L}\big(\textstyle\sum_{k=1}^n \xi_k\big), \mathrm{Po}(\mu)\Big) + d_2\big(\mathrm{Po}(\mu), \mathrm{Po}(\lambda)\big) \tag{7.8}$$

with $\mu(B) = \sum_{k=1}^n \mathbb{P}[|\xi_k| \ge 1, S_k \in B]$ for any Borel set $B \subset E'$.

The second summand is estimated by Theorem 4.3.D as

$$d_2\big(\mathrm{Po}(\mu), \mathrm{Po}(\lambda)\big) \le \min\Big(1, \frac{1.65}{\sqrt{|\mu|}}, \frac{1.65}{\sqrt{|\lambda|}}\Big)\Big||\mu| - |\lambda|\Big| + \big(1 - e^{-\min(|\mu|,|\lambda|)}\big)\, d_W\Big(\frac{\mu}{|\mu|}, \frac{\lambda}{|\lambda|}\Big). \tag{7.9}$$

We have

$$|\mu| = n\big(1 - e^{-\sigma_1 h}\big), \quad |\lambda| = \sigma_1 n h\,,$$

and

$$\frac{\mu}{|\mu|}\big([0,t] \times C\big) = \frac{1 - e^{-\sigma_1 h t}}{1 - e^{-\sigma_1 h}}\Big(\frac{\beta_1}{\sigma_1}\delta_0 + \frac{\alpha_1}{\sigma_1}\delta_1\Big)(C),$$
$$\frac{\lambda}{|\lambda|}\big([0,t] \times C\big) = t\Big(\frac{\beta_1}{\sigma_1}\delta_0 + \frac{\alpha_1}{\sigma_1}\delta_1\Big)(C),$$

for any $t \in [0,1]$ and $C \subset \{0,1\}$. The Wasserstein term in (7.9) can easily be estimated by noting that, since $\mu/|\mu|$ and $\lambda/|\lambda|$ are product measures that put both the same mass on $[0,1] \times \{0\}$, as well as on $[0,1] \times \{1\}$,

$$d_W\Big(\frac{\mu}{|\mu|}, \frac{\lambda}{|\lambda|}\Big) = d_W\Big(\frac{\mu}{|\mu|}(\cdot \times \{0,1\}), \frac{\lambda}{|\lambda|}(\cdot \times \{0,1\})\Big),$$

where the underlying distances are d_0 on the left hand side, and the Euclidean distance on the right hand side. Using then the fact that, for real-valued random variables X and Y, the Wasserstein distance between their distributions can be represented as

$$d_W(\mathscr{L}(X), \mathscr{L}(Y)) = \int_{-\infty}^{\infty} \big|\mathbb{P}[X \le x] - \mathbb{P}[Y \le x]\big|\, dx$$

(compare Equation (3.11)), yields

$$d_W\Big(\frac{\mu}{|\mu|},\frac{\lambda}{|\lambda|}\Big)=\int_0^1\Big|\frac{1-e^{-\sigma_1 ht}}{1-e^{-\sigma_1 h}}-t\Big|\,dt\le\frac{(\sigma_1 h)^2}{4(1-e^{-\sigma_1 h})}.$$

Thus, we obtain in Inequality (7.9)

$$d_2\big(\mathrm{Po}(\mu),\mathrm{Po}(\lambda)\big)\le\Big(1\wedge\frac{1.65}{\sqrt{\sigma_1 nh}}\Big)\frac{\sigma_1^2}{2}nh^2+\frac{1-e^{-\sigma_1 nh}}{1-e^{-\sigma_1 h}}\frac{\sigma_1^2h^2}{4}\le\frac{3\sigma_1^2}{4}nh^2. \tag{7.10}$$

The first summand in Inequality (7.8) is suited for the application of Theorem 7.2.A. For the terms in the upper bound of that theorem, we obtain

$$\begin{aligned}
q_k&=1-e^{-\sigma_1 h}\le\sigma_1 h,\ \text{and}\\
q_k'&=\mathbb{P}[T_k+T_k'\le h]\\
&=\int_0^h\Big(\mathbb{P}[T_k'\le h-t\,|\,U_k=1]\,\mathbb{P}[U_k=1]\\
&\qquad+\mathbb{P}[T_k'\le h-t\,|\,U_k=0]\,\mathbb{P}[U_k=0]\Big)\sigma_1e^{-\sigma_1 t}\,dt\\
&=\int_0^h\Big(\big(1-e^{-\sigma_2(h-t)}\big)\alpha_1+\big(1-e^{-\sigma_0(h-t)}\big)\beta_1\Big)e^{-\sigma_1 t}\,dt\\
&\le\frac{1}{2}(\alpha_1\sigma_2+\beta_1\sigma_0)h^2.
\end{aligned}\tag{7.11}$$

Choosing an arbitrary $r\ge 0$, and setting $\Theta_k^s:=\big\{l\in\mathbb{N}\setminus\{k\};\,|z_l-z_k|\le r\big\}$ and $\Theta_k^w:=\big\{l\in\mathbb{N};\,|z_l-z_k|>r\big\}$ for the neighborhoods of strongly and weakly dependent processes, respectively, yields furthermore

$$\begin{aligned}
\mathbb{P}\big[|\xi_k|\ge 1,|\xi_l|\ge 1\big]&\le h\phi(h)\ \text{ and}\\
\mathbb{E}\Big|\mathbb{P}\big[|\xi_k|\ge 1\,|\,(\xi_l)_{l\in\Theta_k^w}\big]-q_k\Big|&\le\psi_1(h,r),
\end{aligned}\tag{7.12}$$

because ξ_l is a measurable function of $\zeta_l|_{[0,h]}$ for every $l\in\mathbb{N}$, and

$$\begin{aligned}
&q_k\,\mathbb{E}\,d_W\Big(\mathscr{L}\big(S_k\;\big|\;|\xi_k|\ge 1\big),\mathscr{L}\big(S_k\;\big|\;|\xi_k|\ge 1,(\xi_l)_{l\in\Theta_k^w}\big)\Big)\\
&\quad\le q_k\,\mathbb{E}\,d_{TV}\Big(\mathscr{L}\big(S_k\;\big|\;|\xi_k|\ge 1\big),\mathscr{L}\big(S_k\;\big|\;|\xi_k|\ge 1,(\xi_l)_{l\in\Theta_k^w}\big)\Big)\\
&\quad\le\mathbb{E}\Big(\sup_{B\in\mathcal{B}_{[0,1]}}\Big|\mathbb{P}\big[S_k\in B,|\xi_k|\ge 1\;\big|\;(\xi_l)_{l\in\Theta_k^w}\big]-\mathbb{P}\big[S_k\in B,|\xi_k|\ge 1\big]\Big|\Big)\\
&\qquad+\mathbb{E}\Big(\sup_{B\in\mathcal{B}_{[0,1]}}\Big|\mathbb{P}\big[S_k\in B\;\big|\;|\xi_k|\ge 1,(\xi_l)_{l\in\Theta_k^w}\big]\Big(\mathbb{P}\big[|\xi_k|\ge 1\big]-\mathbb{P}\big[|\xi_k|\ge 1\;\big|\;(\xi_l)_{l\in\Theta_k^w}\big]\Big)\Big|\Big)\\
&\quad\le\psi_2(h,r)+\psi_1(h,r)
\end{aligned}\tag{7.13}$$

for the same reason. The expectations above are well-defined, because the suprema can all be replaced by suprema over countable sets, for example in lines 2 to 4 by the suprema

over all finite unions of intervals with endpoints in $\mathbb{Q}\cap[0,1]$ (which can be shown by using an elementary approximation property for finite measures). This fact also justifies the inequality between the supremum and the essential supremum used for the last line.

Combining the estimates from Inequalities (7.10) to (7.13), we obtain

$$\begin{aligned} &d_2\Big(\mathscr{L}\big(\textstyle\sum_{k=1}^n \xi_k^{(h)}\big), \mathrm{Po}(\lambda_{n,h})\Big) \\ &\qquad \le \left(\frac{3\sigma_1^2}{4} + \frac{1}{2}(\alpha_1\sigma_2 + \beta_1\sigma_0) + M_2(\mu)\sigma_1^2\right) nh^2 \\ &\qquad\quad + M_2(\mu)\overline{c}\sigma_1^2 r^D nh^2 + M_2(\mu)\overline{c}r^D nh\phi(h) \\ &\qquad\quad + \big(M_1(\mu) + 2M_2(\mu)\big) n\psi_1(h,r) + M_2(\mu) n\psi_2(h,r), \end{aligned} \tag{7.14}$$

where

$$\begin{aligned} M_1(\mu) &= 1 \wedge \frac{1.65}{\sqrt{n(1-e^{-\sigma_1 h})}} \le 1 \wedge \frac{5}{2\sqrt{\sigma_1 nh}} \quad \text{and} \\ M_2(\mu) &= 1 \wedge \left[\frac{2}{n(1-e^{-\sigma_1 h})}\Big(1 + 2\log^+\Big(\frac{n(1-e^{-\sigma_1 h})}{2}\Big)\Big)\right] \\ &\le 1 \wedge \left[\frac{4}{\sigma_1 nh}\Big(1 + 2\log^+\Big(\frac{\sigma_1 nh}{2}\Big)\Big)\right] \end{aligned}$$

for $h \le 1/\sigma_1$. Since $r \ge 0$ was arbitrary, this yields the required upper bound. □

Remark 7.3.C (A sketch for the model with randomly positioned animals). It might be desirable to model also the positions of the animals as random. Then an upper bound can be calculated in a similar fashion as above, but with a few important differences: Typically, one wants to drop Condition (7.4) in this situation and work with $\varrho(\mathbb{B}(z_k, r))-1$ directly, where ϱ is now a point process and z_k its k-th point (in a suitable enumeration). Accordingly, the index sets Θ_k^s and Θ_k^w defined after Inequality (7.11) are now random. It is no problem to adapt Theorem 7.2.A so that it comprises random index sets: Θ_k^w appears only via the random variable $\Xi_k^w = \sum_{l\in\Theta_k^w} I_l\delta_{S_l}$ in the proof of Theorem 7.2.A, and it is easily seen that the few properties of Ξ_k^w we used remain unchanged for random Θ_k^w. The set Θ_k^s on the other hand, appears only as a summation set in the estimation of the first and the fourth summand in Inequality (7.2). There, the only difference is that the summation and the expectation cannot be exchanged. In total, we get the same bound in Theorem 7.2.A for the case of random index sets as for the case with deterministic index sets, except for the third summand, which is, in the random case,

$$M_2(\mu)\sum_{k=1}^{\infty} \mathbb{E}\left(\sum_{l\in\Theta_k^s}\big(q_k 1_{\{|\xi_l|\ge 1\}} + 1_{\{|\xi_k|\ge 1, |\xi_l|\ge 1\}}\big)\right). \tag{7.15}$$

Thus a very similar upper bound for the d_2-distance in Proposition 7.3.A can be obtained if ϱ is random, but we have to replace Condition (7.5) by suitable conditions that control the term (7.15).

Appendix

A.1 A bound for Poisson tail probabilities

The following Lemma is used several times in this thesis. Its first inequality is noted without proof in Barbour, Holst, and Janson (1992) as Proposition A.2.3(ii).

Lemma A.1.A. *Let $Z \sim \mathrm{Po}(\lambda)$, where $\lambda > 0$. Then*

$$\mathbb{P}[Z \geq m] \leq \frac{m+1}{m+1-\lambda}\mathbb{P}[Z = m] < \frac{1}{\sqrt{2\pi m}}\frac{m+1}{m+1-\lambda}\Big(\frac{\lambda}{m}\Big)^m e^{m-\lambda}$$

for any $m \in \mathbb{N}$ with $m > \lambda - 1$.

Proof. We have

$$\mathbb{P}[Z \geq m] = \sum_{k=m}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda} \leq \frac{\lambda^m}{m!}e^{-\lambda}\sum_{k=0}^{\infty}\Big(\frac{\lambda}{m+1}\Big)^k = \frac{m+1}{m+1-\lambda}\mathbb{P}[Z = m].$$

Quantitative forms of the DeMoivre-Stirling Theorem, such as the one in Theorem 5.44 of Stromberg (1981), yield

$$m! > \sqrt{2\pi m}e^{-m}m^m \quad \text{for any } m \in \mathbb{N},$$

so that

$$\frac{m+1}{m+1-\lambda}\mathbb{P}[Z = m] < \frac{1}{\sqrt{2\pi m}}\frac{m+1}{m+1-\lambda}\Big(\frac{\lambda}{m}\Big)^m e^{m-\lambda}.$$

□

A.2 Some topological results

A.2.1 The lcscH E

We list here a few properties of the general lcscH E introduced in Section 2.1 which are needed in the main text. Proofs are given by combining several more general results from topology. All the numbers of cited results refer to Dugundji (1966). The spaces appearing in these results are always tacitly assumed to be Hausdorff.

Proposition A.2.A. *Every lcscH E is Polish; that is, there is a metric d inducing the topology on E such that (E, d) is a complete, separable metric space.*

Proof. Separability does not depend on the chosen metric. By Theorem VIII.7.3 (in Dugundji (1966)), every second countable (Hausdorff) space is separable. By Theorem XI.6.4, every locally compact space is completely regular, hence regular; metrizability then follows by Urysohn's Theorem (Corollary IX.9.2: in second countable spaces regularity is equivalent to metrizability). By Corollary XIV.2.4, every locally compact metric space has an equivalent metric that makes the space complete. Hence the proposition is shown. □

Proposition A.2.B. *Every lcscH E is σ-compact.*

Proof. By Theorem VIII.6.3 every second countable Hausdorff space is Lindelöf (i.e. every open covering contains a countable subcovering). For every $s \in E$, let $U(s)$ be an open neighborhood such that $\overline{U(s)}$ is compact (such a neighborhood exists, because E is locally compact). Since $\{U(s); s \in E\}$ is an open covering of E, there must be a countable subcovering $\{U(s_i); i \in \mathbb{N}\}$. Thus $\overline{U(s_i)}$ are compact and $\bigcup_{i\in\mathbb{N}} \overline{U(s_i)} = E$. □

A.2.2 Spaces of measures on E

The *vague topology* on $\mathfrak{M}$ is the topology which has as a basis the family of all sets of the form

$$\bigcap_{i=1}^{n} \left\{ \mu \in \mathfrak{M};\, a_i < \int f_i \, d\mu < b_i \right\} \tag{A.1}$$

with $n \in \mathbb{N}$, $f_i \in \mathfrak{F}_{cc}$, and $a_i, b_i \in \mathbb{R}$ for $i \in \{1, \ldots, n\}$. The same holds true for the *vague topology* on $\mathfrak{N}$, provided we replace $\mathfrak{M}$ by $\mathfrak{N}$ in (A.1). We denote these topologies by $\mathcal{T}_{\mathfrak{M}}$ and $\mathcal{T}_{\mathfrak{N}}$, respectively. The convergence concept induced by these topologies is called *vague convergence*. If $\mu, \mu_1, \mu_2, \ldots$ are measures in $\mathfrak{M}$ (or $\mathfrak{N}$), we have $\mu_n \xrightarrow{v} \mu$, that is μ_n converges vaguely to μ, if and only if $\int f d\mu_n \to \int f d\mu$ for every $f \in \mathfrak{F}_{cc}$. From Kallenberg (1986) we quote the following two results about the vague topologies.

Proposition A.2.C. $(\mathfrak{M}, \mathcal{T}_{\mathfrak{M}})$ *and* $(\mathfrak{N}, \mathcal{T}_{\mathfrak{N}})$ *are Polish.*

Proof. See Kallenberg (1986), Result 15.7.7. □

Proposition A.2.D. *For the standard σ-algebras $\mathcal{M}$ on $\mathfrak{M}$ and $\mathcal{N}$ on $\mathfrak{N}$ given in Subsection 2.2.1 of the main text, that is*

$$\mathcal{M} := \sigma\big([\mathfrak{M} \ni \mu \mapsto \mu(B)];\, B \in \mathcal{B} \text{ bounded}\big) \text{ and}$$
$$\mathcal{N} := \sigma\big([\mathfrak{N} \ni \mu \mapsto \mu(B)];\, B \in \mathcal{B} \text{ bounded}\big),$$

we have

$$\mathcal{M} = \sigma(\mathcal{T}_{\mathfrak{M}}) \text{ and } \mathcal{N} = \sigma(\mathcal{T}_{\mathfrak{N}}).$$

Proof. See Kallenberg (1986), Lemma 4.1. □

A.3 Locally evaluable random fields

In Section 6.1, the notion of a locally evaluable random field was introduced. Essentially, any measurable random field which owes its measurability to some local feature (such as continuity of paths) has this property. In what follows we show local evaluability for two important classes of random fields. We first give the corresponding definitions.

Definition. Let $p : \mathbb{R}_+^D \to [0,1]$ be a function, and $\pi := \{\pi(\cdot, s); s \in \mathbb{R}_+^D\}$ a $[0,1]$-valued random field on $\mathbb{R}_+^D$.

(i) Let $A \subset \mathbb{R}_+^D$ be a closed convex cone, that is a closed convex set for which $v \in A$ implies $\alpha v \in A$ for every $\alpha \in \mathbb{R}_+$. We call p *continuous from A at a point $s \in \mathbb{R}_+^D$* if $\lim_{n\to\infty} p(s_n) = p(s)$ for any sequence (s_n) in $s + A$ with $s_n \to s$ $(n \to \infty)$. We call p *continuous from A* if it is continuous from A at every point.

(ii) We say p is *lower semicontinuous at a point $s \in \mathbb{R}_+^D$* if $\liminf_{n\to\infty} p(s_n) \geq p(s)$ for every sequence (s_n) in $\mathbb{R}_+^D$ with $s_n \to s$ $(n \to \infty)$. We say p is *upper semicontinuous at s* if $-p$ is lower semicontinuous at s. Furthermore p is called *lower [resp. upper] semicontinuous* if it is lower [resp. upper] semicontinuous at every point.

(iii) Let $\mathcal{C}$ be a class of subsets of $\mathbb{R}$. We call π *separable for $\mathcal{C}$* if there exists a countable set $\Sigma \subset \mathbb{R}_+^D$, and a fixed set $N_0 \in \mathcal{A}$ with $\mathbb{P}(N_0) = 0$, such that for any set $C \in \mathcal{C}$ and for any rectangle R that is open in $\mathbb{R}_+^D$, we have

$$\big\{\omega \in \Omega;\, \pi(\omega, s) \in C, s \in R\big\} \,\Delta\, \big\{\omega \in \Omega;\, \pi(\omega, s) \in C, s \in R \cap \Sigma\big\} \subset N_0,$$

where Δ denotes the symmetric difference operation for sets. In this case we call Σ a *separant* for π. We call π *fully separable for $\mathcal{C}$* and accordingly Σ a *full separant* for π if the above property holds with $N_0 = \emptyset$. Accordingly, every separable random field π can be made fully separable by adjustment on a set of probability zero. Note that such an adjustment does not change the distribution of the thinning ξ_π.

Proposition A.3.A. *The random field $\pi := \{\pi(\cdot, s); s \in \mathbb{R}_+^D\}$ is locally evaluable under each of the following conditions:*

(i) There is a closed convex cone $A \subset \mathbb{R}_+^D$ with non-empty interior, such that the paths $\pi(\omega, \cdot)$ are all continuous from A;

(ii) the paths $\pi(\omega, \cdot)$ are all lower semicontinuous, and π is fully separable with respect to the class $\{(y, \infty);\, y \in \mathbb{R}\}$;

(iii) the paths $\pi(\omega, \cdot)$ are all upper semicontinuous, and π is fully separable with respect to the class $\{(-\infty, y);\, y \in \mathbb{R}\}$.

Remark A.3.B. A special case of a random field with all upper semicontinuous paths is the indicator of a random closed set (RACS) Ξ; see Subsection 2.3.1 for the definition. It is straightforward to see that such an indicator is separable with respect to the class $\{(-\infty, y);\, y \in \mathbb{R}\}$ if and only if Ξ is a separable RACS, meaning that there is a countable dense subset Σ of $\mathbb{R}_+^D$ such that $\Xi = \overline{\Xi \cap \Sigma}$ a.s. Hence Proposition A.3.A guarantees us, if $\pi(\omega, s) := 1_{\Xi(\omega)}(s)$ for a separable RACS Ξ, that by adjusting π on a null set, we obtain a locally evaluable random field. For more details on the relationship between indicator random fields and RACS see Matheron (1975), Section 2-4.

Proof of Proposition A.3.A. (i) Take D linearly independent vectors $v_1, v_2, \dots, v_D \in A$ of unit length, and let $V := \{\sum_{i=1}^D \alpha_i v_i;\ \alpha_i \in (0,1]\}$ be the half-open parallelepiped spanned by these vectors. Define the lattice $L^{(n)} := \{2^{-n}\sum_{i=1}^D k_i v_i;\ k_i \in \mathbb{Z}\}$, and denote by $V_v^{(n)}$ the shifted copy of $2^{-n}V$ with the maximal corner in $v \in L^{(n)}$ (i.e. of all the corners $2^{-n}\sum_{i=1}^D k_i v_i$ of $V_v^{(n)}$ the corner v maximizes the coefficients $k_1, \dots, k_D$). Note that $\{V_v^{(n)};\ v \in L^{(n)}\}$ is, for any n, a partition of $\mathbb{R}^D$. We may now, for any open set $R \subset \mathbb{R}_+^D$, approximate π_R^* (see Section 6.1 for the notation) by the random fields π_n given as

$$\pi_n(\omega, s) := \sum_{v \in L^{(n)} \cap R} \pi(\omega, v) 1_{V_v^{(n)}}(s)$$

for every $\omega \in \Omega$ and every $s \in R$. It is easy to see that π_n is $(\sigma(\pi|_R) \otimes \mathcal{B}_R)$-measurable. Furthermore, π_n converges pointwise to π_R^* for $n \to \infty$, because for any (ω, s), the sequence $(\pi_n(\omega, s))_n$ has (with the possible exception of the first finitely many elements) the same values as $(\pi(\omega, s_n))_n$, where s_n are defined via the condition $s \in V_{s_n}^{(n)}$ for every $n \in \mathbb{N}$, such that $s_n \to s$ from the cone A. Hence, π_R^* is $\big(\sigma(\pi|_R) \otimes \mathcal{B}_R\big)$-measurable, which completes the proof of part (i).

(ii) We show that $(\pi_R^*)^{-1}\big((y,\infty)\big) \in \sigma(\pi|_R) \otimes \mathcal{B}_R$ for any rectangle R that is open in $\mathbb{R}_+^D$ and for any $y \in [0,1)$. Since π is pathwise lower continuous, we get for all $\omega \in \Omega$ that $\big(\pi(\omega,\cdot)\big)^{-1}\big((y,\infty)\big) \cap R$ is open. Let Σ be a full separant for π, and set $\mathcal{G} := \{(a,b) \subset R;\ a, b \in \mathbb{Q}^D, a < b\}$. We then can write

$$(\pi_R^*)^{-1}\big((y,\infty)\big) = \bigcup_{G \in \mathcal{G}} \bigcap_{s \in G \cap \Sigma} \big(\pi(\cdot, s)\big)^{-1}\big((y,\infty)\big) \times G$$

for any $y \in [0,1)$, which lies in $\sigma(\pi|_R) \otimes \mathcal{B}_R$, because the union and the intersection are countable.

(iii) Apply (ii) to $1 - \pi$.

□

A.4 Proofs left out in the main text

A.4.1 Abstract d_2-estimate for the Cox approximation of thinnings (proof of Proposition 6.2.B)

We use the notation from Sections 6.1 and 6.2. Let $n \in \mathbb{N}$ be fixed, and let $\sum_{i=1}^{\xi(\mathbb{R}_+^D)} \delta_{S_i}$ be a representation of ξ with $\sigma(\xi)$-measurable random elements $S_1, S_2, \dots$ numbered in such a way that all the points in $\kappa_n^{-1}(C)$ come first. Let $(X_i)_{i\in\mathbb{N}}$, $(Y_i)_{i\in\mathbb{N}}$ be sequences of random variables such that, given ξ and π_n, the X_i are independent indicators with expectations $\pi_n(S_i)$, and the Y_i are independent and $\mathrm{Po}(\pi_n(S_i))$-distributed (i.e. the X_i are retention decisions as usual, and the Y_i are "Poissonized" variants of retention decisions). We have

$$\begin{aligned} &d_2\big(\mathscr{L}(\xi_{\pi_n}\kappa_n^{-1}|_C), \mathrm{Cox}(\Lambda|_C)\big) \\ &\qquad \le d_2\big(\mathscr{L}(\xi_{\pi_n}\kappa_n^{-1}|_C), \mathrm{Cox}(\Lambda_n|_C)\big) + d_2\big(\mathrm{Cox}(\Lambda_n|_C), \mathrm{Cox}(\Lambda|_C)\big). \end{aligned}$$

Choosing an arbitrary $\mathrm{Cox}(\Lambda_n)$-distributed point process η_n, the first term is estimated as

$$\begin{aligned} &d_2\big(\mathscr{L}(\xi_{\pi_n}\kappa_n^{-1}|_C), \mathrm{Cox}(\Lambda_n|_C)\big) \\ &\qquad \le \mathbb{E}\bigg(\sup_{f\in\mathfrak{F}_2}\Big|\mathbb{E}\big(f(\xi_{\pi_n}\kappa_n^{-1}|_C) - f(\eta_n|_C) \;\big|\; \xi, \pi_n\big)\Big|\bigg) \\ &\qquad = \mathbb{E}\bigg(d_2\Big(\mathscr{L}\big(\textstyle\sum_{i=1}^{\xi\kappa_n^{-1}(C)} X_i\delta_{\kappa_n(S_i)} \;\big|\; \xi, \pi_n\big), \mathscr{L}\big(\sum_{i=1}^{\xi\kappa_n^{-1}(C)} Y_i\delta_{\kappa_n(S_i)} \;\big|\; \xi, \pi_n\big)\Big)\bigg) \\ &\qquad \le 6.5\,\mathbb{E}\bigg(\sup_{s\in\kappa_n^{-1}(C)} \pi_n(s)\bigg), \end{aligned} \tag{A.2}$$

where for the last line we used the fact that, given ξ and π_n, each of the sequences (X_i) and (Y_i) is independent, so that Theorem 4.3.C can be applied.

For the second term, we just use an upper bound for the d_2-distance between two general Cox processes. Let Λ and M be arbitrary finite random measures on C, and write $\Phi(\lambda,\mu) := d_2\big(\mathrm{Po}(\lambda), \mathrm{Po}(\mu)\big)$ for finite measures λ and μ on C. Then, by conditioning on Λ and M,

$$d_2\big(\mathrm{Cox}(\Lambda), \mathrm{Cox}(\mathrm{M})\big) \le \mathbb{E}\big(\Phi(\Lambda, \mathrm{M})\big), \tag{A.3}$$

and by Theorem 4.3.D,

$$\begin{aligned} \Phi(\lambda,\mu) \;\le\; &\big(1 - e^{-\min(\lambda(C),\mu(C))}\big)\, d_W\bigg(\frac{\lambda}{\lambda(C)}, \frac{\mu}{\mu(C)}\bigg) \\ &+ \min\bigg(1, \frac{1.65}{\sqrt{\lambda(C)}}, \frac{1.65}{\sqrt{\mu(C)}}\bigg)\Big|\lambda(C) - \mu(C)\Big|. \end{aligned} \tag{A.4}$$

Combining of Estimates (A.2), (A.3), and (A.4) yields the desired result.

A.4.2 Thinnings are determined locally (result needed for Inequalities (6.13) and (6.15))

We show here that "local conditioning" is enough to determine the "local distribution" of the thinnings, a result which is required for the estimation of the $e_{\boldsymbol{kr}}$-terms in both Section 6.3 and Section 6.4. For any set $A \subset \mathbb{R}_+^D$ write

$$\mathcal{N}(A) := \sigma\big(\big\{\{\varrho \in \mathfrak{N}; \varrho(\tilde{A}) = l\}\,;\; \tilde{A} \in \mathcal{B}_A, l \in \mathbb{Z}_+\big\}\big).$$

We then obtain

Lemma A.4.A. *Suppose that ξ is a point process on $\mathbb{R}_+^D$.*

(i) **For the π-thinning:**
Let π be a locally evaluable $[0,1]$-valued random field on $\mathbb{R}_+^D$. Then we have for any bounded open set $A \subset \mathbb{R}_+^D$, and for any $N \in \mathcal{N}(A)$, that

$$\mathbb{P}\big[\xi_\pi \in N \;\big|\; \xi, \pi\big] = \mathbb{P}\big[\xi_\pi \in N \;\big|\; \xi|_A, \pi|_A\big] \quad a.s.$$

(ii) **For the Q-thinning:**
Let $(Q_u)_{u\in\mathbb{N}}$ be an admissible sequence of retention kernels which satisfy Assumption 3 from Section 6.4. Then we have for any bounded measurable set $A \subset \mathbb{R}^D_+$ and for any $N \in \mathcal{N}(A)$, that

$$\mathbb{P}\big[\xi_Q \in N \;\big|\; \xi\big] = \mathbb{P}\big[\xi_Q \in N \;\big|\; \xi|_{A^{[\tilde{t}_n]}}\big] \quad a.s.,$$

where $A^{[\tilde{t}_n]} := \big\{s \in \mathbb{R}^D_+;\, |s - s'| \le \tilde{t}_n \text{ for some } s' \in A\big\}$.

Proof. We only prove (i), since the proof of (ii) is very similar, requiring only a few and very obvious changes in notation. Let $A \subset \mathbb{R}^D_+$ be bounded and open, and let $N \in \mathcal{N}(A)$. We show that $P_{(\xi,\pi)}(N)$, the version of $\mathbb{P}\big[\xi_\pi \in N \;\big|\; \xi, \pi\big]$ used for the π-thinning definition in Section 6.1, is $\sigma(\xi|_A, \pi|_A)$-measurable, whence the statement follows.

Let $k \in \mathbb{N}$, $B_1, \dots B_k \in \mathcal{B}_A$ be pairwise disjoint, and $l_1, \dots, l_k \in \mathbb{Z}_+$. Choose a representation $\sum_{i=1}^{\xi(A)} \delta_{S'_i}$ of $\xi|_A$ with $\sigma(\xi|_A)$-measurable random elements $S'_1, S'_2, \dots$, numbered "B_i-wise" (first all points in B_1, then all points in B_2, and so on, until B_k). Any such representation can easily be extended to a representation $\sum_{i=1}^{\xi(\mathbb{R}^D_+)} \delta_{S_i}$ of ξ with $\sigma(\xi)$-measurable random elements $S_1, S_2, \dots$, numbered "B_i-wise". We then have for $N = \{\varrho \in \mathfrak{N};\, \varrho(B_1) = l_1, \dots, \varrho(B_k) = l_k\}$, setting $V_j := \sum_{i=1}^{j} \xi(B_i)$ for $j \in \{0, 1, \dots, k\}$ and $t_0 := 0$,

$$P_{(\xi,\pi)}(N) \;=\; \sum_{\substack{t_1,\dots,t_k=0\\ t_1\le\dots\le t_k}}^{\infty} \mathrm{I}\big[V_j = t_j\,;\; 1 \le j \le k\big] \cdot \sum_{e_1,\dots,e_{t_k}\in\{0,1\}} \mathrm{I}\big[\textstyle\sum_{i=t_{j-1}+1}^{t_j} e_i = l_j\,;\; 1 \le j \le k\big] \prod_{i=1}^{t_k} \pi(S_i)^{e_i}\big(1 - \pi(S_i)\big)^{1-e_i},$$

which by $\pi(S_i) \equiv \pi|_A(S'_i)$, and the local evaluability of π, is obviously $\sigma(\xi|_A, \pi|_A)$-measurable (note that the measurability property required for local evaluability extends easily to arbitrary bounded open sets, instead of only bounded open rectangles). Since any set of the form $N = \{\varrho \in \mathfrak{N};\, \varrho(\widetilde{B}_1) = l_1, \dots, \varrho(\widetilde{B}_k) = l_k\}$ with arbitrary $\widetilde{B}_1, \dots \widetilde{B}_k \in \mathcal{B}_A$ can be written as a finite disjoint union of sets of the form $\{\varrho \in \mathfrak{N};\, \varrho(B_1) = l_1, \dots, \varrho(B_k) = l_k\}$ with pairwise disjoint $B_1, \dots B_k \in \mathcal{B}_A$, we obtain immediately the $\sigma(\xi|_A, \pi|_A)$-measurability of $P_{(\xi,\pi)}(N)$ for sets N of this new form. The measurability for general N is now obtained by a standard extension argument. □

A.4.3 Simplification for the retention kernels in the Q-thinning (proof of Remark 6.4.A)

We use the notation from Section 6.4. Let $(f_i)_{i\in\mathbb{N}}$ be a sequence of $\mathcal{N}$-$\mathcal{B}^D_+$-measurable functions, and suppose that $Q_u\big((f_1(\sigma), \dots, f_u(\sigma); \sigma), A\big)$ is defined for every $\sigma \in \mathfrak{N}$, every $A \subset \{0,1\}^u$, and every $u \in \mathbb{N}$, subject to the same conditions as in Remark 6.4.A. For $u \in \mathbb{N}$ and $(s_1, s_2, \dots, s_u; \sigma) \in E_u$, choose pairwise different $i_1, i_2, \dots, i_u \in \mathbb{N}$ with $f_{i_j}(\sigma) = s_j$, $1 \le j \le u$, and let $\bar{u} := \max\{i_j;\, 1 \le j \le u\}$. By Properties (a) and (b) it is clear that if an extension $\widetilde{Q}_u$ of Q_u with the required features exists, it must have the form

$$\widetilde{Q}_u\big((s_1, \dots, s_u; \sigma), A\big) = Q_{\bar{u}}\Big(\big(f_1(\sigma), \dots, f_{\bar{u}}(\sigma); \sigma\big), T_2\big(A \times \{0,1\}^{\bar{u}-u}\big)\Big),$$

where $A \subset \{0,1\}^u$ and $T_2 : \{0,1\}^{\bar{u}} \to \{0,1\}^{\bar{u}}$ is the transformation that belongs (in the sense of Property (a) from the Q-thinning definition) to an arbitrary permutation τ on $\{1,2,\ldots,\bar{u}\}$ with $\tau(i_j) := j$ for $1 \le j \le u$. It is easy to see that the $\widetilde{Q}_u$ are well-defined, extend the Q_u, and satisfy Properties (a) and (b) from the Q-thinning definition. In order to obtain the measurability in the first argument, note that for any $A \subset \{0,1\}^u$ and any measurable $B \subset [0,1]$,

$$\big(\widetilde{Q}_u(\cdot, A)\big)^{-1}(B) = \bigcup_{\substack{i_1,\ldots,i_u \in \mathbb{N} \\ i_j \neq i_k \text{ for } j \neq k}} \Big\{ \big(f_{i_1}(\sigma), \ldots, f_{i_u}(\sigma); \sigma\big) \,;\, \sigma \in \varphi_{i_1,\ldots,i_u}^{-1}(B) \Big\},$$

where $\varphi_{i_1,\ldots,i_u}(\sigma) := \widetilde{Q}_u\big((f_{i_1}(\sigma), \ldots, f_{i_u}(\sigma); \sigma), A\big)$ for every $\sigma \in \mathfrak{N}$. By the definition of the $\widetilde{Q}_u$ and the measurability requirement on the Q_u, it follows that $\varphi_{i_1,\ldots,i_u}^{-1}(B) \in \mathcal{N}$. Hence the above set is a countable union of graphs of $\mathcal{N}$-$(\mathcal{B}_+^D)^u$-measurable functions each of which is intersected by a $(\mathcal{B}_+^D)^u \otimes \mathcal{N}$-measurable set. Thus $\big(\widetilde{Q}_u(\cdot, A)\big)^{-1}(B) \in (\mathcal{B}_+^D)^u \otimes \mathcal{N}$, which completes the proof.

A.5 Details for the point pattern data presented in Chapter 1

In Chapter 1, a selection of point pattern data is presented (see Figure 1.1). We provide here some additional information about the nature of these data sets, in the order in which they are displayed in Figure 1.1. Except for the first example, all data sets are taken from the spatstat package* of the statistical programming environment R†.

Epicenters of earthquakes in Southern California

Displayed are the epicenters (201 in total) of all local seismic events of magnitude above 4.5 in the database of the Southern Californian Earthquake Data Center (SCEDC) that took place from 1985 to 2004 in the selected rectangular region. This region ranges from 122 to 114 western longitude and from 32 to 37 northern latitude, which roughly corresponds to Southern California and its closer neighborhood. Figure A.5.1 shows the point pattern given in Chapter 1 on a simple map‡.

Apparently strong clustering occurs in the point pattern; accumulations of points are typically found along major geological faults. The well-known San Andreas fault, for example, runs roughly diagonally from the top left to the bottom right of the picture.

*Adrian Baddeley, Rolf Turner, with contributions by Marie-Colette van Lieshout and 17 others (2005). spatstat: spatial point pattern analysis, model-fitting and simulation. R package version 1.6-8. http://www.maths.uwa.edu.au/∼adrian/spatstat.html

†R Development Core Team (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

‡drawn with GMT, the Generic Mapping Tools. http://gmt.soest.hawaii.edu

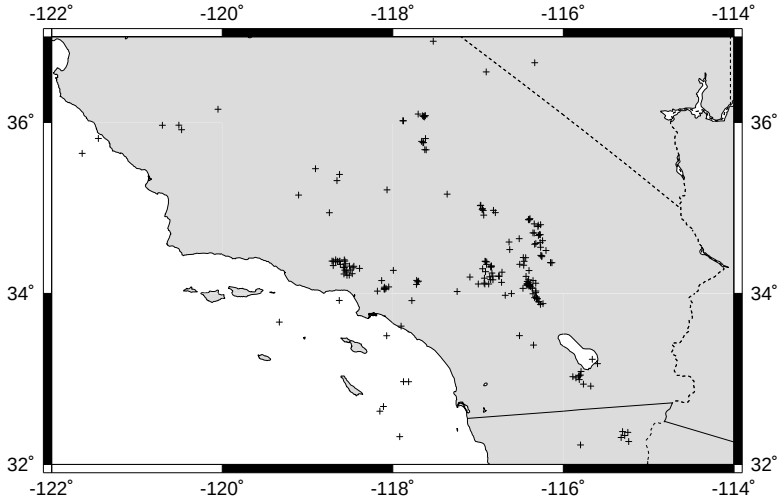

FIGURE A.5.1: THE EARTHQUAKE POINT PATTERN WITH SOME GEOGRAPHICAL INFORMATION.

Seedlings and saplings of California redwood trees

Displayed are 62 seedlings and saplings of California redwood trees (Sequoia sempervirens) on a square area of sidelength 23 meters, approximately. The dataset was extracted by Ripley (1977) from a dataset given on a larger area by Strauss (1975). Again we see strong clustering, which in this case can be attributed to the tendency for the seedlings to be positioned around stumps of redwood trees.

Centers of biological cells

This plot shows a pattern that gives a completely different visual impression. The centers of 42 biological cells as observed under a microscope are shown. The point pattern exhibits apparent spatial regularity, which is probably due to the fact that the cells themselves are of considerable size compared to the scale of the plot. The data were recorded by Francis Crick and Brian Ripley (see Ripley (1977)).

Nests of two species of ants in northern Greece

Displayed are the locations of the nests of two species of ants, Messor wasmanni ($\triangle$) and Cataglyphis bicolor ($\bigcirc$), in an irregular region of about 130 meters diameter in northern Greece. The data were recorded by R.D. Harkness. There are 68 Messor and 29 Cataglyphis nests. Questions of interests concern both intra- and inter-species relationships. The Messor are harvester ants which live mainly on seeds, the Cataglyphis are foragers which eat dead insects (and other arthropods), among other things dead Messor ants that have been killed by some other source. Biological hypotheses based on these behavioral patterns include that there is spatial inhibition among the Messor nests (which eyeball observation of the data seems to confirm), and that there is a tendency for the Cataglyphis nests to be situated close to the Messor nests. For an extensive analysis of the data see Baddeley and Turner (2005).

Saplings of Japanese black pines

The final plot shows 65 saplings of Japanese black pines (Pinus thunbergii) in a natural forest in Chiba prefecture, Japan. The displayed area has an original size of 5.7 by 5.7 meters. The data were first published in Numata (1961). They are often used as an example of an "interaction free" point pattern which may very well be modeled by a homogeneous Poisson process. The superficial similarity of this pattern with the patterns in Figure 1.2 is a manifestation of this.

Bibliography

Baddeley, A., Turner, R. (2005). Modelling spatial point patterns in R. In: *Case studies in spatial point process modeling*, edited by A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, Lecture Notes in Statistics 185. Springer, New York, 23–76.

Banys, R. (1980). On superpositions of random measures and point processes. In: *Mathematical statistics and probability theory (Proc. Sixth Internat. Conf., Wisła, 1978)*, Lecture Notes in Statistics 2. Springer, New York, 26–37.

Barbour, A.D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* Special Vol. **25A** 175–184.

Barbour, A.D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322.

Barbour, A.D., Brown, T.C. (1992). Stein's method and point process approximation. *Stochastic Process. Appl.* **43** 9–31.

Barbour, A.D., Chen, L.H.Y., Loh W.-L. (1992). Compound Poisson approximation for nonnegative random variables via Stein's method. *Ann. Probab.* **20** 1843–1866.

Barbour, A.D., Chryssaphinou, O. (2001), Compound Poisson approximation: a user's guide. *Ann. Appl. Probab.* **11** 964–1002.

Barbour, A.D., Holst, L., Janson, S. (1992). *Poisson approximation.* Oxford University Press, Oxford.

Barbour, A.D., Månsson, M. (2002). Compound Poisson process approximation. *Ann. Probab.* **30** 1492–1537.

Böker, F., Serfozo, R. (1983). Ordered thinnings of point processes and random measures. *Stochastic Process. Appl.* **15** 113–132.

Borovkov, A.A. (1996). Asymptotic expansions for functionals of dilation of point processes. *J. Appl. Probab.* **33** 573–591.

Brown, T.C. (1979). Position dependent and stochastic thinning of point processes. *Stochastic Process. Appl.* **9** 189–193.

Brown, T.C., Weinberg, G.V., Xia, A. (2000). Removing logarithms from Poisson process error bounds. *Stochastic Process. Appl.* **87** 149–165.

Brown, T.C., Xia, A. (1995a). On metrics in point process approximation. *Stochastics Stochastics Rep.* **52** 247–263.

Brown, T.C., Xia, A. (1995b)., On Stein-Chen factors for Poisson approximation. *Statist. Probab. Lett.* **23** 327–332.

Brown, T.C., Xia, A. (2001). Stein's method and birth-death processes. *Ann. Probab.* **29** 1373–1403.

Chen, L.H.Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545.

Chen, L.H.Y., Shao, Q. (2005). Stein's method for normal approximation. In: *An introduction to Stein's method*, edited by A.D. Barbour and L.H.Y. Chen, IMS Lecture Notes Series Vol. 4. World Scientific, Singapore, 1–59.

Cohn, D.L. (1980). *Measure theory.* Birkhäuser, Boston.

Conway, J.H., Sloane, N.J.A. (1999). *Sphere packings, lattices and groups.* Springer, New York.

Daley, D.J. (1974). Various concepts of orderliness for point processes. In: *Stochastic Geometry*, ed. E.F. Harding and D.G. Kendall. Wiley, New York, 148–161.

Daley, D.J., Vere-Jones, D. (1988). *An Introduction to the theory of point processes.* Springer, New York.

Diggle, P.J. (1983). *Statistical analysis of spatial point patterns.* Academic Press, London.

Doukhan, P. (1994). *Mixing: Properties and examples*, Lecture Notes in Statistics 85. Springer, New York.

Dudley, R.M. (1989). *Real analysis and probability.* Wadsworth & Brooks/Cole, Pacific Grove, CA.

Dugundji, J. (1966). *Topology.* Allyn and Bacon, Boston.

Ehm, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11** 7–16.

Ellis, S.P. (1986). A limit theorem for spatial point processes. *Adv. in Appl. Probab.* **18** 646–659.

Ellis, S.P. (1991). Density estimation for point processes. *Stochastic Process. Appl.* **39** 345–358.

Erhardsson, T. (2005). Stein's method for Poisson and compound Poisson approximation. In: *An introduction to Stein's method*, edited by A.D. Barbour and L.H.Y. Chen, IMS Lecture Notes Series Vol. 4. World Scientific, Singapore, 61–113.

Feller, W. (1968). *An introduction to probability theory and its applications, vol. I.* Third edition. Wiley, New York.

Gibbs, A.L., Su, F.E. (2002). On choosing and bounding probability metrics. *International Statistical Review* **70**, no. 3, 419–435.

Goldman, J.R. (1967). Stochastic point processes: limit theorems. *Ann. Math. Statist.* **38** 771–779.

Goldstein, L., Rinott, Y. (1996), Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1–17.

Grigelionis, B. (1963). On the convergence of sums of random step processes to a Poisson process. *Theory Probab. Appl.* **8** 177–182.

Jagers, P. (1972). On the weak convergence of superpositions of point processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **22** 1–7.

Jagers, P., Lindvall, T. (1974). Thinning and rare events in point processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **28** 89–98.

Kabatiansky, G.A., Levenshtein, V.I. (1978). Bounds for packings on a sphere and in space. *Problems of Information Transmission* **1** 1–17.

Kallenberg, O. (1975). Limits of compound and thinned point processes. *J. Appl. Probability* **12** 269–278.

Kallenberg, O. (1986). *Random measures*, 4th ed. Akademie-Verlag, Berlin; Academic Press, London.

Kallenberg, O. (2002). *Foundations of modern probability*, 2nd ed. Springer, New York.

Khinchin, A.Y. (1955). *Mathematical methods in the theory of queueing (in Russian).* Trudy Mat. Inst. Steklov v. 49, Izdat. Akad. Nauk SSSR, Moscow.
Translated: Griffin, London, 1960.

Luk, H.M. (1994). *Stein's method for the gamma distribution and related statistical applications.* Ph.D. thesis, University of Southern California, Los Angeles.

Matheron, G. (1975). *Random sets and integral geometry.* Wiley, New York.

Matthes, K., Kerstan, J., Mecke, J. (1978). *Infinitely divisible point processes.* Wiley, Chichester.

Møller, J., Waagepetersen, R.P. (2004). *Statistical inference and simulation for spatial point processes.* Chapman & Hall/CRC, Boca Raton, FL.

Musin, O.R. (2003). The problem of the twenty-five spheres. *Russian Math. Surveys* **58** 794–795.

Neveu, J. (1965). *Mathematical foundations of the calculus of probability.* Holden-Day, San Francisco.

Numata, M. (1961). Forest vegetation in the vicinity of Choshi. Coastal flora and vegetation at Choshi, Chiba Prefecture IV (in Japanese). *Bull. Choshi Marine Lab. Chiba Univ.* **3** 28–48.

Palm, C. (1943). *Intensitätsschwankungen im Fernsprechverkehr.* Ericsson Technics no. 44.

Papadimitriou, C.H., Steiglitz, K. (1982). *Combinatorial optimization: algorithms and complexity.* Prentice-Hall, Englewood Cliffs, NJ.

Rachev, S.T. (1991). *Probability metrics and the stability of stochastic models.* Wiley, Chichester.

Reinert, G. (1994). *A weak law of large numbers for empirical measures via Stein's method, and applications.* Ph.D. thesis, Universität Zürich.

Reinert, G. (2005). Three general approaches to Stein's method. In: *An introduction to Stein's method*, edited by A.D. Barbour and L.H.Y. Chen, IMS Lecture Notes Series Vol. 4. World Scientific, Singapore, 183–221.

Rényi, A. (1957). A characterization of Poisson processes (in Hungarian with summaries in English and Russian). *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **1** 519–527. Translated in *Selected papers of Alfréd Rényi*, Vol. 1, ed. Pál Turán, Akadémiai Kiadó, Budapest, 1976.

Ripley, B.D. (1977). Modelling spatial patterns (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 172–212.

Ripley, B.D. (1981). *Spatial statistics.* Wiley, New York.

Schneider, R., Weil, W. (1992). *Integralgeometrie.* Teubner, Stuttgart.

Schneider, R., Weil, W. (2000). *Stochastische Geometrie.* Teubner, Stuttgart.

Schuhmacher, D. (2005a). Upper bounds for spatial point process approximations. *Ann. Appl. Probab.* **15** 615–51.

Schuhmacher, D. (2005b). Distance estimates for Poisson process approximations of dependent thinnings. *Electron. J. Probab.* **10** 165–201.

Schuhmacher, D. (2005c). Distance estimates for dependent superpositions of point processes. *Stochastic Process. Appl.* **115** 1819–1837.

Serfozo, R.F. (1977). Compositions, inverses and thinnings of random measures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **37** 253–265.

Serfozo, R.F. (1984a). Rarefactions of compound point processes. *J. Appl. Probability* **21** 710–719.

Serfozo, R. (1984b). Thinning of cluster processes: convergence of sums of thinned point processes. *Math. Oper. Res.* **9** 522–33.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* Chapman and Hall, London.

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 2.* Univ. California Press, Berkeley, CA, 583–602.

Stein, C. (1986). *Approximate computation of expectations*, Institute of Mathematical Statistics Lecture Notes 7. Hayward, CA.

Stoyan, D., Kendall, W.S., Mecke, J. (1987). *Stochastic geometry and its applications.* Wiley, Chichester.

Strauss, D.J. (1975). A model for clustering. *Biometrika* **62** 467–475.

Stromberg, K.R. (1981). *Introduction to classical real analysis.* Wadsworth, Belmont, CA.

van Lieshout, M.N.M., Baddeley, A.J. (1996). A nonparametric measure of spatial interaction in point patterns. *Statist. Neerlandica* **50** 344–361.

Xia, A. (2005). Stein's method and Poisson process approximation. In: *An introduction to Stein's method*, edited by A.D. Barbour and L.H.Y. Chen, IMS Lecture Notes Series Vol. 4. World Scientific, Singapore, 115–181.

Yannaros, N. (1991). Poisson approximation for random sums of Bernoulli random variables. *Statist. Probab. Lett.* **11** 161–165.

Zeger, K., Gersho, A. (1994). Number of nearest neighbors in a Euclidean code. *IEEE Trans. Inform. Theory* **40** 1647–1649.

Curriculum vitae

Name:	SCHUHMACHER
Vornamen:	Dominic Markus
Geburtsdatum:	22.04.1974
Heimatort:	Kloten ZH

Ausbildung

1981–1987:	Primarschule Hinterwiden, Kloten
1987–1993:	Kantonsschule Zürcher Unterland, Bülach
September 1993:	Matura Typus B
1994–2000:	Studium an der Universität Zürich Hauptfach: Mathematik Nebenfächer: Informatik, Volkswirtschaftslehre
November 2000:	Diplom in Mathematik Titel der Diplomarbeit: *Distanzabschätzungen für Poissonprozeß-Approximationen im $\mathbb{R}^2$ unter Verwendung einer verallgemeinerten Stein-Chen-Methode* (ausgeführt unter der Leitung von Prof. Dr. Andrew Barbour)
seit November 2000:	Assistent/Doktorand am Institut für Mathematik, Universität Zürich

Publikationen während des Promotionsstudiums

Februar 2005:	Upper bounds for spatial point process approximations. *Ann. Appl. Probab.* **15** 615–51
Februar 2005:	Distance estimates for Poisson process approximations of dependent thinnings. *Electron. J. Probab.* **10** 165–201
November 2005:	Distance estimates for dependent superpositions of point processes. *Stochastic Process. Appl.* **115** 1819–1837

Vorträge während des Promotionsstudiums

19. September 2001:	60-minütiger Vortrag an der School of Mathematical Sciences der Monash University in Melbourne, Australien
5. März 2002:	30-minütiger Vortrag im Rahmen des *Séminaire de printemps* des *Troisième cycle romand de statistique et probabilités appliquées* in Champéry (VS)
22. März 2002:	20-minütiger Vortrag im Rahmen der Magdeburger Stochastiktage, Deutschland
19. August 2003:	30-minütiger Vortrag im Rahmen von *Stein's method and applications: a program in honor of Charles Stein* am Institute for Mathematical Sciences der National University of Singapore
14. September 2004:	15-minütiger Vortrag im Rahmen der *CIME summer school on stochastic geometry* in Martina Franca, Italien